Sleep, 17(2):146–159

A Field Study of Sleep Disturbance: Effects of Aircraft Noise and Other Factors on 5,742 Nights of Actimetrically Monitored Sleep in a Large Subject Sample

*J. A. Horne, *F. L. Pankhurst, *L. A. Reyner, †K. Hume and §I. D. Diamond

**Department of Human Sciences, Loughborough University, U.K.;*
†Department of Biological Sciences, Manchester Metropolitan University, U.K. and
§Department of Social Statistics, University of Southampton, U.K.

Summary: This field study assessed the effects of nighttime aircraft noise on actimetrically measured sleep in 400 people (211 women and 189 men; 20–70 years of age; one per household) habitually living at eight sites adjacent to four U.K. airports, with different levels of night flying. Subjects wore wrist-actimeters for 15 nights and completed morning sleep logs. A sample of 178 nights of sleep electroencephalograms (EEGs) were recorded synchronously with actigrams. The EEG was used to develop filters for the raw actigrams, in order to: (1) estimate sleep onset and (2) compare actigrams with aircraft noise events (ANEs). Actigrams, filtered to detect the onset of discrete movements, were able to detect 88% of all EEG-determined periods of interim wakefulness of >15 seconds and periods of movement time of >10 seconds. The main findings were: (1) actimetry and self-reports showed that only a minority of ANEs affected sleep, and, for most of our subjects, that domestic and idiosyncratic factors had much greater effects; (2) despite large between-site variations in ANEs, the difference between sites in overall sleep disturbance was not significant; (3) there was a diminished actimetric response to ANEs in the first hour of sleep and, apparently, also in the last hour of sleep; and (4) men had significantly more discrete movements than women and were more likely to respond to ANEs. **Key Words:** Noise—Sleep disturbance—Actimetry—Sleep EEG—Subjective responses.

The recent comprehensive review by Pearsons et al. (1) on the effects of noise on sleep, not only pointed to a shortage of appropriate field studies in comparison with laboratory investigations, but also showed that people seem to be much more sensitive to equivalent noises in the laboratory than in the field. This can be seen clearly in Fig. 1, which is based on studies covered in the review (1). The percentage of subjects awakened is plotted against indoor noise event levels.

In a field study requiring objective assessment of sleep, the financial and other costs of all-night sleep encephalogram (EEG) recordings allow only a limited number of subjects to be monitored at once. Body movements are a good index of sleep quality and especially reflect arousals (even small arousals) during sleep (2–4). Thus, a simpler, relatively inexpensive and feasible alternative to EEGs is to employ wrist-worn actimeters, so that many devices can be used at once. Slow limb movements (e.g. passive body movement due to mattress displacement by a moving bed partner) are unlikely to be registered by these devices as the acceleration is usually less than the threshold set in the actimeter. Poor or disturbed sleep is reflected by increased actimetrically measured movements (5). There is also good evidence that noise-disturbed sleep is usually accompanied by increased limb movements (6,7).

The assessment of sleep by automatic recording of body motility predates the EEG measurement of sleep, and many impressive studies were undertaken even in the 1920s [see Kleitman (3)]. However, it was not until more recent times that direct comparisons were also made with the EEG [for review see Tryon (4)]. One of the earliest evaluations of wrist-worn actimeters against the sleep EEG was performed by Mullaney et al. (8), who analyzed 102 subject nights on a 1-minute epoch basis. There was a 94.5% agreement between the two methods for the estimates of sleep onset and waking.

Accepted for publication December 1993.

Address correspondence and reprint requests to J. A. Horne, Department of Human Sciences, Loughborough University, Leicestershire LE11 3TU, U.K.

More detailed analyses of actimeter and EEG data in 10 subjects (8), showed a 96% agreement between the two methods in the detection of periods of wakefulness (arousals) within sleep. A later study by the same group (9), over 20 nights using smaller, more sophisticated actimeters, reported a 94% agreement between the EEG and actimetry in identifying sleep and wakefulness episodes throughout the night. Recent studies also report good correlations between EEG and actimetry with regard to the detection of sleep and wakefulness: better than 85% agreement (10–12); normal subjects, better than 90% agreement (12) and various patient groups, 78–90% agreement (12). An automated scoring system to distinguish sleep from wakefulness with actimetry (13) showed the method to be correct on 88% of occasions, when compared with the EEG. Other studies assessing the application of actimetry to insomnia (14,15) have also endorsed the method.

We examined the effects of aircraft noise and other disturbances on actimetrically determined sleep of people living under flight paths at four U.K. airports (Heathrow, Gatwick, Stansted and Manchester). The study was part of a more extensive investigation (16) that also used questionnaire surveys. Our particular interest concerns disturbance within sleep rather than impairments to sleep onset and premature awakening at the end of sleep.

METHOD

Sites

Two sites were selected from each of four airports (Heathrow, Gatwick, Stansted and Manchester), making eight sites in all (designated A–H). Each site contained a sufficient number of dwellings within a small area so that the outdoor level of any single aircraft noise event (ANE) for all the houses concerned would be similar, ideally to within 3 dB. Each site was close to aircraft flight paths but away from other major sources of nighttime noise (e.g. motorways, railways). An overflying aircraft would traverse this zone in a few seconds. The sites were chosen to give a wide range of nighttime ANEs, in terms of both noise levels and number of ANEs.

All ANE levels described in this study relate to outdoor rather than indoor noise. It was not feasible to record indoor noise levels from every bedroom on the scale that would have been required. Such ANE levels would depend on several factors, especially window type and whether windows were open or closed (the latter was recorded for each subject night). Relative to outdoors, ANEs heard indoors would be 10–15 dB lower if the windows were open, 15–25 dB lower with single-glazed closed windows, and up to 35 dB lower with double-glazed closed windows.

The airports differed considerably in the level of nighttime air traffic. Whereas site E averaged around 20 ANEs per night, site H averaged between one and two; the latter was the nearest equivalent to a control site. It was not possible to select control sites away from airports. Representative types of subject with respect to age and gender were selected from each site. Subjects had their sleep monitored for 15 consecutive nights.

Disturbance within sleep

An initial task was to validate the use of actimeters for measuring sleep disturbance. The first step was to establish an acceptable and feasible definition of "sleep disturbance", based on EEG criteria, yet applicable to actimeters and appropriate within the context of this study. The literature is unclear about exact definitions of this term and its synonyms (e.g. sleep disruption), and interpretations range from a transient change of sleep to a lighter stage, to a full awakening lasting for >1 minute. Major reference manuals (e.g. 17,18) give little specific guidance. The views of several internationally known sleep experts were gathered, and although there were clear differences of opinion, all agreed that an EEG-measurable period of interim wakefulness is indicative of sleep disturbance.

Transitory wakefulness (W) in the EEG signal, lasting a few seconds, is difficult to identify, particularly when it is obscured by movement artifact [i.e. movement time (MT), which is often assumed to be a brief awakening]. The EEG criterion we adopted for assessing the ability of actimeters to detect sleep disturbance was any period of wakefulness of >15 seconds or MT of >10 seconds. Most briefer body movements are detected by actimeters (4), and one would expect more actimeter responses than periods of MT or W registered by the EEG. However, our EEG analyses have not been analyzed down to epochs <10 seconds.

EEG data

Sleep EEGs were recorded on a sample of 48 subjects (approximately equal numbers of each sex in the three age ranges: 20–34 years, 35–49 years and 50–70 years), using Oxford Medical Systems Medilog 9200 recorders. The clocks in these devices are accurate to within ±5 seconds per 24 hours and were checked every night. Each subject was recorded for 4 nights, and out of the possible 192 nights, 178 satisfactory recordings were obtained (93%). The EEGs were "blind" scored (18) in 30-second epochs in a hybrid manner, both visually and by the Oxford Sleep Stager (the latter allows a 2-second classification). Any epoch containing >15 seconds of W or >10 seconds of MT was scored as

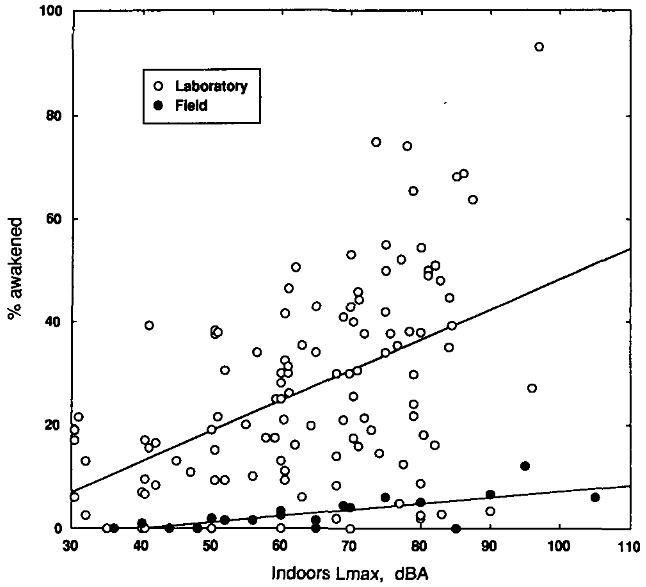


FIG. 1. Probability of noise-induced awakenings in relation to amplitude of noise for all the field and laboratory studies reviewed by Pearsons et al. (1). Note: (i) All noise levels are for indoors, and (ii) subjects in the field, who are usually more adapted to the noise, are less likely to be awakened by noise than subjects who are exposed to similar noise levels in the laboratory.

such, regardless of what the remaining EEG epoch indicated. Hypnograms were transferred to a data base and processed to identify the W or MT epochs that were tagged as "hypnoblips" (see Fig. 2); any series of consecutive EEG epochs of MT, W or both would be tagged only for the first epoch. These data were then synchronized with the corresponding actigrams.

Actimeters

The Swiss-type wrist-worn actimeters (Gaehwiler Electronics-Hombrechtikon) were used. Data (actigrams) were compared with time-logged sound recordings of ANEs to see whether there were concomitant actimeter responses. Fifty subjects wearing synchronized actimeters were monitored concurrently at each site, and the summed group response to each ANE was examined. Our initial rationale was that because body movements during sleep are essentially random, summation of all the actigrams across subjects should reduce the random activity to a residual near-constant background level. A movement common to several subjects, such as that in response to an ANE, would

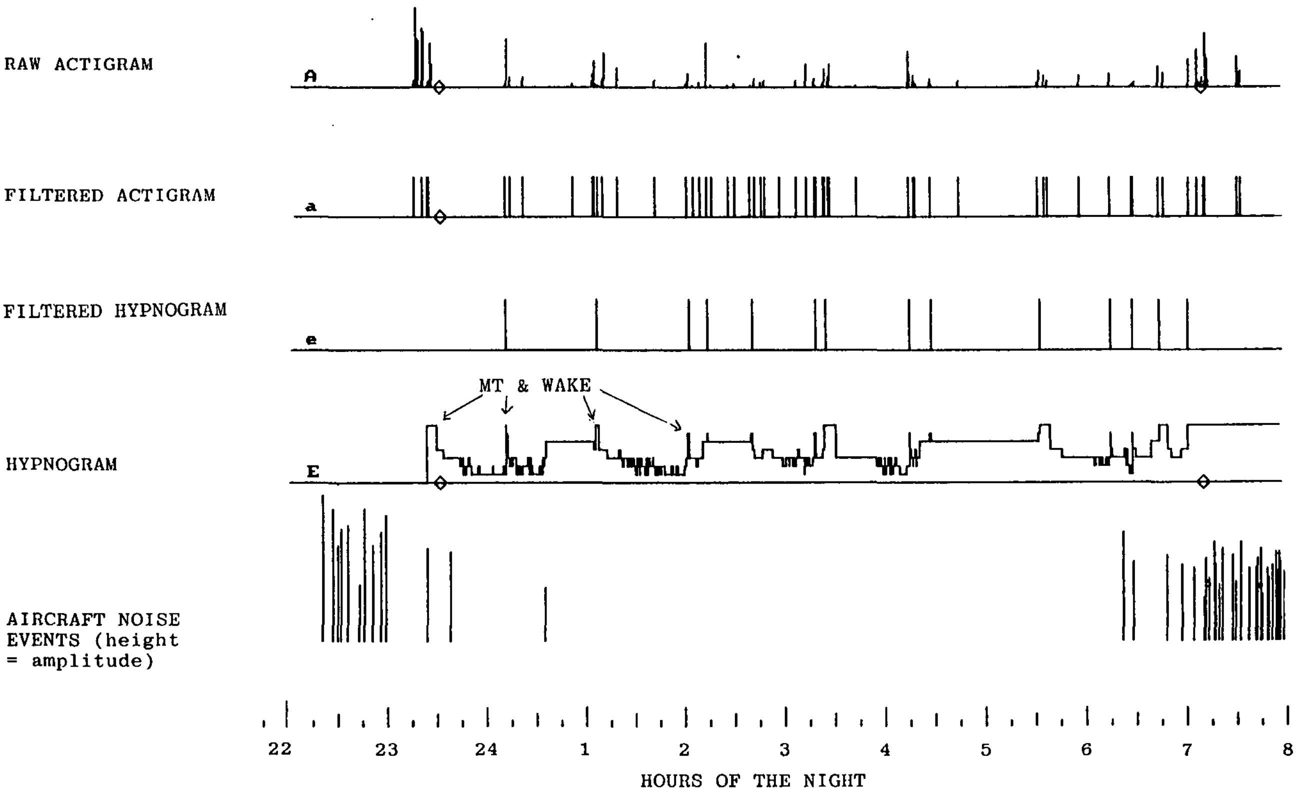


FIG. 2. Top to bottom: "A" = a raw actigram recorded in 30-second epochs; "a" = filtered version of this actigram using a 1, 1, 1 filter; "e" = filtered version of the hypnogram (below) detecting the onset of MT and wakefulness periods; "E" = an EEG-based hypnogram scored in 30-second epochs (synchronized with the actigram to within 5 seconds); Bottom = aircraft noise events (height of line = L_{max} dB). The approximately 900 actigram epochs have been much compressed laterally, and what appear to be thicker actiblips in trace "a" are single actiblips separated by a small gap that is not visible. Diamond markers to the left of "A" and "a" are the estimated sleep onset, using the sleep onset algorithm (see Appendix). Diamond markers to the right are the wake-up times entered by the subject in the morning log.

TABLE 1. *Actimeter vs. EEG results from the EEG subsample of 48 subjects (six per site) recorded for a total of 178 nights. X/H% gives the percentage of EEG-determined sleep disturbances correctly identified by actiblips*

Site	A	X	H	E	X/A%	X/H%	H/E%	A/E%	H/A%
Heathrow (A)	848	322	355	17,300	38.0	90.7	2.05	4.90	41.86
Gatwick (B)	952	327	392	19,183	34.3	83.4	2.04	4.96	41.18
Heathrow (C)	689	294	334	17,540	42.7	88.0	1.90	3.93	48.48
Gatwick (D)	461	146	169	9,240	31.7	86.4	1.83	4.99	36.66
Manchester (E)	761	258	280	17,821	33.9	92.1	1.57	4.27	36.79
Manchester (F)	875	281	312	16,833	32.1	90.1	1.85	5.20	35.66
Stansted (G)	916	342	385	20,170	37.3	88.8	1.91	4.54	42.03
Stansted (H)	957	256	303	17,556	26.8	84.5	1.73	5.45	31.86
All	6,459	2,226	2,530	135,643	34.5	88.0	1.86	4.76	39.17

Notation: A = number of actigraph blips, X = number of coincident blips (±1 epoch), H = number of hypnogram blips, and E = total number of epochs.

evoke a summed response above this background, rather like an EEG-evoked response, which is proportional to the number of subjects responding. As there were clear individual differences in the magnitude of movement responses registered, a simple summation of these data across subjects led to distortions and standardization was needed.

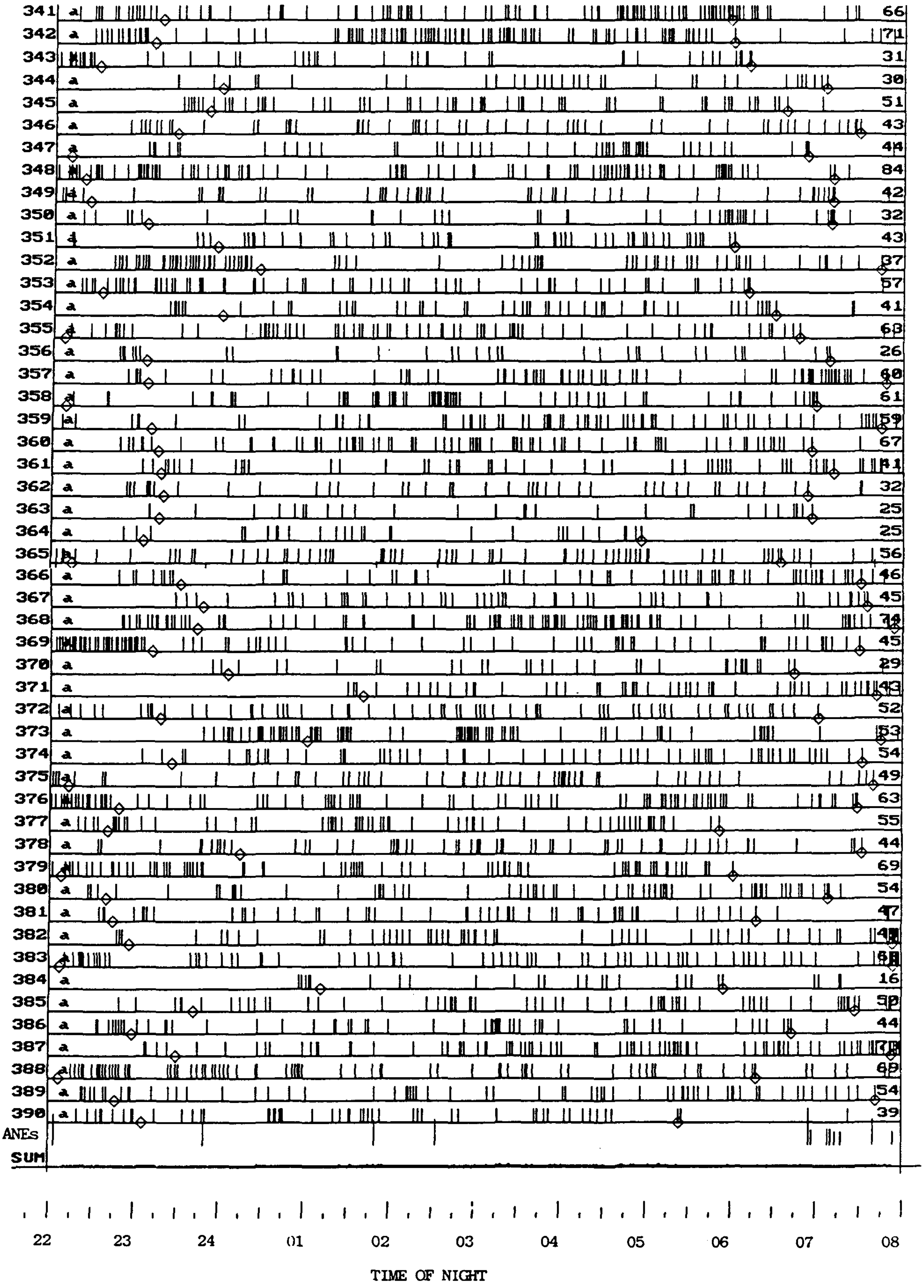
Actimeters were calibrated frequently against a master clock (National Physical Laboratory, U.K.) to within ±5 seconds of real time. They accumulated data in 30-second epochs, giving about 850 epochs from a typical night's sleep (some 42,500 epochs when 50 subjects were monitored concurrently). These data were standardized by means of computer analysis using a program "ACCORD", that transformed each 30-second epoch into a binary on/off. The form this filter took and its validation came from the 178 nights of sleep EEGs that were synchronized to within ±5 seconds of the corresponding actigrams. The filter that best fitted the hypnoblips registered "on" for an actigram epoch containing one or more movement counts, following at least one epoch of less than one unit (i.e. nil) of movement activity. These three numbers were used to define the filter (i.e. 1, 1, 1); this filter is one of the simplest and most sensitive, as it looks for the onset of any small limb movement. The "on" responses which signify the onset of discrete movements, are referred to as "actiblips" (see Fig. 2). All other epochs were scored as zero.

Table 1 gives the correspondence for ±1 epoch between hypnoblips and actiblips for all sites. About 1.9% of all hypnogram (30-second) epochs contained a hypnoblip, compared with about 4.7% for actiblips in the actigrams (for the EEG subjects a typical 7.25-hour night contained about 16 hypnoblips and about 41 actiblips; the latter was slightly higher for the whole subject population). Had the EEG epochs been scored by the usual minimum 15 seconds of W or MT per 30-second epoch, then there would have been fewer apparent episodes of W or MT.

For all the sites, 88% of all hypnoblips matched actiblips to within ±1 epoch. For the approximately 800 remaining epochs of the night, both the EEG and actigraphs concurred that there was a zero score. Over a typical night, and on a 30-second epoch basis, the binary analyses of the EEGs and actigrams agreed on the presence or absence of a disturbance onset for 97% of all epochs. The approximately 24 remaining actiblips per night that were not matched by a hypnoblip were, on more detailed EEG analyses, often associated with minor EEG events such as very short periods of MT/possible awakenings (<10 seconds), shifts to stage 1 and K complexes [associations between body movements and such events have been described by Muzet et al. (2)]. Of course, not all of the latter two events are associated with an actiblip. With regard to the analysis of the EEG for MT/possible awakenings of <10 seconds, it should be noted that visual scoring of the EEG "blind" to detect these events is difficult and unreliable, particularly if there is no accompanying movement artifact. Thus, the hit rate of actiblips on very short duration awakenings is not known.

Analysis

Figure 3 shows 50 subjects' filtered actigrams synchronized for 1 night; actiblips appear as vertical bars. The diamond markers at the beginning of each trace indicate the estimated sleep onset (SO, see Appendix). The subjects' reported final waking up times, taken from morning sleep logs, are identified by a second diamond marker at the end of the trace. The thicker actiblips in Fig. 3 are closely packed single blips, and it should be noted that each actigram consists of about 800 epochs. The ANEs are shown at the bottom as bars; their heights give an indication of the outdoor sound level. The summated data for all subjects ("sum") are shown at the bottom of the figure and are shown magnified in Fig. 4. The middle trace of Fig. 4 gives the actual summation, and the upper trace shows an average, based on the number of subjects actually asleep. ANEs are shown on the lower trace. In Figs. 3 and 4


341 a 66
342 a 71
343 31
344 a 30
345 a 51
346 a 43
347 a 44
348 84
349 42
350 a 32
351 a 43
352 a 37
353 a 57
354 a 41
355 a 68
356 a 26
357 a 60
358 a 61
359 a 59
360 a 67
361 a 41
362 a 32
363 a 25
364 a 25
365 56
366 a 46
367 a 45
368 a 74
369 45
370 a 29
371 a 43
372 a 52
373 a 53
374 a 54
375 a 49
376 63
377 a 55
378 a 44
379 69
380 a 54
381 a 47
382 a 48
383 a 68
384 a 16
385 a 50
386 a 44
387 a 70
388 a 69
389 a 54
390 a 39
ANEs
SUM
22
23
24
01
02
03
04
05
06
07
08
TIME OF NIGHT

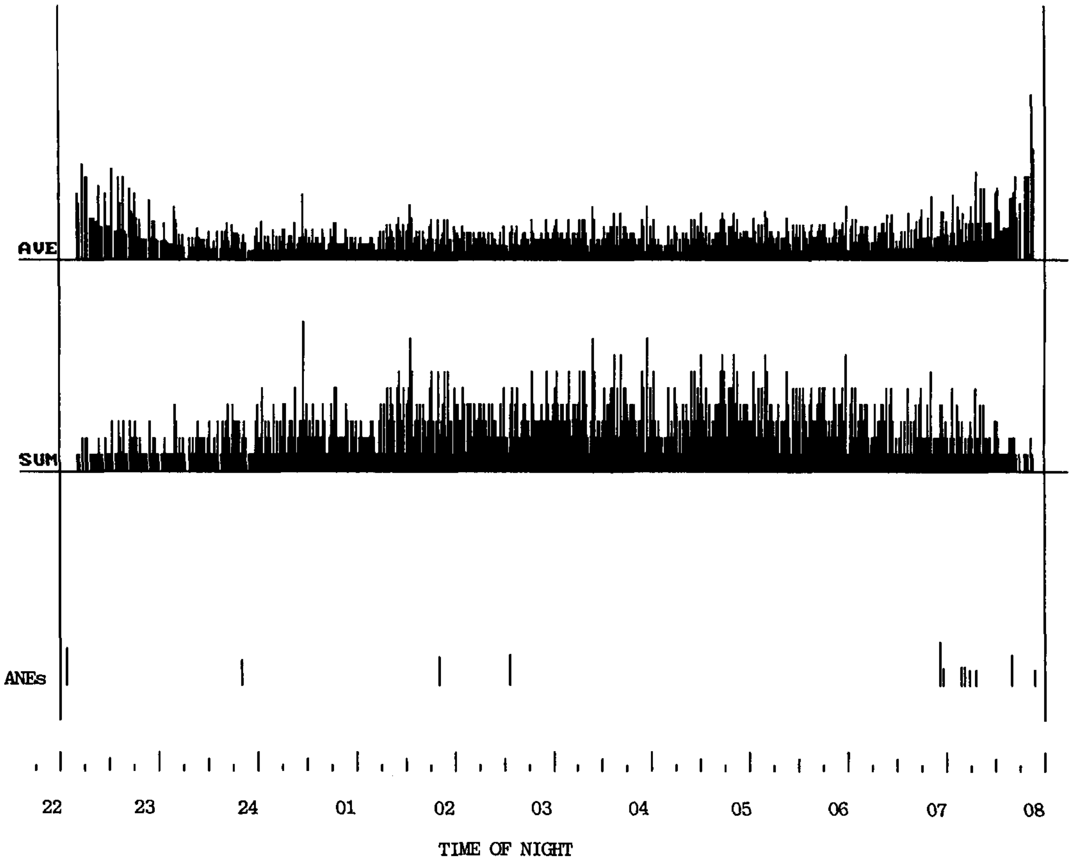


FIG. 4. Middle: summated actigrams (actiblipped) for all subjects (in 30-second epochs) from Fig. 3. Top: Epoch-by-epoch averages based on the number of subjects actually asleep at the time. Bottom: aircraft noise events (height of these bars indicates L_{max}).

the time of night starts at 2200 hours and ends at 0800 hours, but at these extremes few of the subjects are asleep and the summation gives a distorted picture (see below); hence the need for an average.

It is clear from Fig. 4 that none of the ANEs are associated with a summated response beyond the background levels. In fact, there were surprisingly few nights when this was so. It might be asked if this apparent lack of response to ANEs indicated insensitivity of the method or too few people responding to ANEs. The latter alternative is strongly indicated, particularly as the EEG data show that actimeter responses are indeed highly sensitive, at least to periods of W or MT >10–15 seconds. This conclusion is supported by the subjective reports. It should be noted that the great majority of ANEs were below 90 dB (outdoors).

A more discerning analysis of the actigrams was therefore necessary. The method adopted looked at each ANE to see whether an actiblip occurred in the corresponding 30-second epoch for all the appropriately sleeping subjects. The likelihood of this was given as a percentage, designated as noise (n). If an actiblip occurred on all occasions, then n = 100%, if on one in 10 occasions, then n = 10%, etc. The occurrence of actiblips in all other epochs (with no corresponding ANEs) was designated as quiet (q), which was calculated on the same basis as n [e.g. q = 5% means that there was a background (non-ANE) actiblip for one in 20 epochs]. Thus, n − q gives the probability of an ANE itself causing the actiblip. For example, n − q = 2% indicates that one in 50 ANEs causes a sleep disturbance. Again, it is emphasized that in calculating n we only used the actigram epoch in which the ANE occurred, rather than one or more successive epochs. When n values for +1 or +2 epochs following an ANE were calculated, these were not significantly different

---

←

FIG. 3. Filtered actigrams synchronized for 50 subjects (nos. 341–390) for 1 night. Numbers to the right of each trace give the total actiblip count. ANEs are shown at the bottom. The diamond marker to the left of each trace is SO as determined by actigram (see Appendix). The diamond marker to the right of each trace is the wake-up time from the sleep log. Bottom: "sum" = summation of all traces (see Fig. 4).

from q (see Results). That is, any sleep disturbance registered on the actigram, and seemingly caused by an ANE, usually occurred within a few seconds of the ANE.

Subjects

Subjects were chosen from an initial door-to-door questionnaire survey compiled by the authors, undertaken on our behalf by Public Attitude Survey, U.K., a specialist survey organization. For each of the eight sites, approximately 200 subjects were interviewed initially (n = 1,638). This survey is described elsewhere (19). Interviewers knocked on doors and continued interviewing until the quota of 200 was obtained. One member per household was interviewed, with the constraint being that the sample was intended to reflect the age and gender distribution of the local population, based on the latest U.K. census data. This meant that for each site there were approximately equal numbers of men and women and an even distribution across the three age ranges: 20–34 years, 35–49 years and 50–70 years. These respondents were not aware of the specific purpose of the study but were told (correctly) that it was a government sponsored project into living conditions. After the interview they were asked if they would be willing to participate in the main study, and if so, undergo a more detailed interview about their sleep (at a later date). The survey questionnaires of the 971 respondents who agreed were then scrutinized to determine which volunteers were

a. available during the specific measurement period,
b. not deaf,
c. not suffering from nighttime pain that seriously disrupted sleep (e.g. severe arthritis and rheumatism), and
d. not taking sleeping tablets or other medications that promote sleep (including significant consumption of alcohol at bedtime).

This last criterion was problematic, as it could be argued that the associated insomnia of these rejected medicated subjects may have been partly the result of aircraft noise, with the only way for the sufferer to obtain satisfactory sleep being from a hypnotic, which may well suppress any noise effect on sleep. Our survey indicated that 4.4% of the respondents were currently on hypnotics and they were excluded. Nonmedicated, self-diagnosed "poor sleepers" were included. Table 2 gives the proportions of those in the survey and in the main study having, "at least some difficulty in getting to sleep", "regularly awoken up once asleep", "at least some difficulty in getting back to sleep", etc. It can be seen here that about one-third of our subjects had some sleep difficulty at least twice per month.

A total of 614 subjects fulfilled criteria a–d and were given a further structured interview, supervised by two of us (F.L.P. and L.A.R.), covering such topics as anxiety, illnesses, worries, medication, smoking, tea and coffee intake, evening exercise, medical or other reason affecting sleep, difficulty going to sleep, quality of sleep,

TABLE 2. *Characteristics based on social survey responses of all subjects initially interviewed and for the subject subsamples used in measurements by: (i) actimetry and (ii) EEG. Numbers are expressed as percent by category*

		Subject group		
Variable	Category	Social survey (n = 1,636)	Actimetry (n = 400)	EEG (n = 48)
Age (years)	20–34	33	37	35
	35–49	29	33	33
	50–70	37	29	33
Sex	Male	50	48	47
	Female	50	52	53
Marital status	Married	71	75	78
	Single	16	14	16
	Separated etc.	12	11	6
Occupational group	ABC1[h]	48	50	47
	C2DE[i]	51	50	53
Years in residence	<5	31	40	33
	≥5	69	60	67
Children	None	63	55	55
	One or more	37	45	45
Window type	Single	64	64	65
	Double	36	36	35
Windows at night	Open	79	80	78
	Shut	20	19	22
Sleeper	Light	42	44	45
	Deep	58	56	55
Bed partner	None	29	25	20
	Never disturbs	48	46	51
	Disturbs	23	29	29
ANGEN[a]	No	71	71	69
	Yes	29	29	31
ANWK[b]	No	80	77	78
	Yes	20	22	22
HEARNT[c]	No	83	81	84
	Yes	17	19	16
COMPLAIN[d]	No	89	91	84
	Yes	11	9	16
DIFFGET[e]	No	60	63	55
	Yes	40	37	45
WOKENREG[f]	No	84	81	84
	Yes	16	19	16
GETBACK[g]	No	68	72	69
	Yes	32	28	31

There were no significant differences between the three groups for any variable. Note that for each of the scales ANWK, DIFFGET, WOKENREG and GETBACK the "yes" values are for all subjects responding more than "twice a month".

[a] ANGEN = very much annoyed by aircraft noise.
[b] ANWK = awakened at night by aircraft noise.
[c] HEARNT = very much annoyed by aircraft noise at night.
[d] COMPLAIN = has made a formal complaint about noise.
[e] DIFFGET = has difficulty getting to sleep.
[f] WOKENREG = regularly woken up once asleep.
[g] GETBACK = has difficulty getting back to sleep once woken.
[h] ABC1 = professional and skilled.
[i] C2DE = semi-skilled and manual.

mood in the morning, nighttime waking, alertness on arising and at bedtime, and sleeping arrangements. These subjects were then asked to volunteer for sleep measurement by (1) actimeter and EEG or (2) actimeter alone: 220 agreed to the former, 227 to the latter. Payment was £5 (about U.S. $8) per actimeter night and a further £15 (about U.S. $25) per sleep-EEG night. Of the 447 subjects, 400 were selected randomly, and the rest put on reserve. At each site, six subjects were chosen for 4 nights of sleep-EEG recording.

The age ranges and gender distributions of subjects selected for actimetry were as follows: 20–34 years (148: 80 women, 68 men); 35–49 years (135: 73 women, 62 men); and 50–70 years (117: 58 women, 59 men). Total number of subjects selected was 400: 211 women, 189 men.

The subgroups were similar in cross section to those of the larger survey sample, as can be seen from Table 2, which also gives the characteristics of the actimetry and EEG subjects. Sleep was also monitored concurrently by actigrams in 46 bed partners for 8 consecutive nights; the outcomes for this group will be reported elsewhere.

On completion of the actimetry, subjects were given a structured debriefing interview based on a questionnaire that was administered by the same researcher as before. It was only then that subjects were informed of the exact nature of the study (i.e. to investigate the effects of aircraft noise on sleep). These subjects also completed personality/anxiety questionnaires. (Details with regard to the more sleep-disturbed subjects will be covered elsewhere.)

For all actimetry nights, subjects completed morning sleep logs covering bedtime, lights out, estimated SO and number of interim awakenings (with details), times of morning awakening and rising, sleep quality, and alertness 15 minutes after awakening. Subjects undergoing actimetry put on their actimeters when getting into bed, and removed them on morning arising. Subject compliance was good, and out of the 6,000 total possible nights that could have been recorded, 5,742 were obtained (96%).

Measurement of ANEs

Measurement of ANEs was performed by the United Kingdom Civil Aviation Authority (16). In summary, at each site two or three sound level monitors were located some distance apart on the site perimeter. The monitors were linked to a central recording and processing system. Only ANEs with sufficiently distant yet intense sources could trigger all the monitors. More localized noise (e.g. road traffic, animals and neighbors) would be registered only by one sensor. In this way ANEs could be differentiated readily. All outdoor noise events exceeding a threshold of 60 dB, whether they be ANEs or from other sources, were logged in terms of sound levels [especially the highest instantaneous sound level recorded during an ANE (L_{max})] and exact time. ANEs were synchronized with the actigrams and hypnograms, for analyses by ACCORD. From the airport logs, all aircraft movements were identified according to landing/taking-off, route and aircraft type. Subjects reported in their morning logs whether their bedroom windows had been open or closed during the night.

Noise levels were measured on the scales of both L_{max} (in dB) and sound exposure level (SEL, the standard measurement of a noise event taking into account both the intensity and duration). As L_{max} and SEL are very highly correlated, we report all ANE levels in terms of L_{max} only.

RESULTS AND DISCUSSION

General findings

Each subject contributed many observations (e.g. 15 nights × 840 epochs = 12,600 epochs of actigram data per subject). Standard statistical analyses assume that each observation is independent of all others, which is usually not the case here. There are no simple solutions, and it is considered acceptable to apply standard multivariate analyses, as we have done. More advanced methods are being used to analyze these data more rigorously, and further details are given elsewhere (16). In the more conventional analyses that follow, the confidence intervals are approximate and should be treated with caution.

Figure 5 is derived from SOs (based on the SO algorithm described in the Appendix) and the self-reported final wake-up times (from the morning sleep logs) for all 5,742 subject nights. Figure 5 also indicates the numbers of subjects presumed to be asleep every 15 minutes (accumulated from the 30-second epochs). Also shown is a histogram (15-minute intervals) of the percentage of the 30-second sleep epochs when there were ANEs, which sleepers could have heard and to which they could have responded. For the 10-hour time window (2200 to 0800 hours) there was a total of 121,534 sleep epochs when an ANE occurred, to which a sleeper could have responded (see "N" in Table 3). Figure 5 shows that there were relatively few ANEs in the middle of the night.

For most subjects, ANEs had little effect on actiblips. The grand average of q (see above) was 5.36% and is based on almost 5 million epochs. The value for n was 6.49% for all ANEs of L_{max} greater than 60 dB. The n − q of 1.13% indicates that, on average, only one in 88 ANEs resulted in an actiblip response, but there are clear individual differences.

These average n and q values are not based on each individual's n, q and n − q, which are then averaged, but are calculated from pooled data based on all actimeter responses summed across all subjects and given as percentages of all sleep epochs summated for all subjects. (Using each subject's own scores, and averaging these across all subjects, gives very similar outcomes.) All values for n are based on the same epoch in which the ANE occurred. The value for n after a lag of one epoch (n.b. one epoch = 30 seconds) after an ANE is 5.43%, and for lags of two, four, six, eight and 10 epochs after an ANE, the values for n are 5.47%, 5.62%, 5.35%, 5.55% and 5.27%, respectively. Although some of these figures are slightly above the mean q of 5.36%, which might suggest small residual effects, it was decided that the substantial sacrifice of degrees of freedom for every lag accommodated was not worthwhile. As it was clear that the majority of actiblip responses to ANEs occurred in the same 30-second epoch as that of the ANE, we confined ourselves to this particular epoch.

Average n and q values for each subject were correlated ($r = 0.422$; df = 399) across the subjects, indicating that people who seem more disturbed in their sleep are more likely to respond to an ANE. However, most of the variance still remains unexplained.

Due to the low incidence of ANEs at one of the airports (sites G and H, see "N" in Table 3) only about 25% of these particular subjects showed two or more responses to all the ANEs over the 15-day period. Although their n values may not be as reliable as those from subjects at noisier sites, the combined n for sites G and H is 6.39%, which is very close to the overall mean. The corresponding combined q value for these two sites is 5.40%, which is based on about one million data epochs and is reliable.

Concerning open/closed bedroom windows, most subjects habitually slept under constant window conditions. For those who varied this from night to night, there was no significant effect on their n, q or n − q values.

Between sites

Because of the large numbers of epochs used to calculate q and the much smaller numbers used for n, values for "a" (i.e. all epochs with an actiblip = n and q combined) are virtually the same as those for q. Site values for n, q and "a" are given in Table 3. Despite large intersite differences in the number of nighttime ANEs, there are no significant intersite differences for "a" or q with either mean values per subject (Table 3) or average trends over the night (see below). Even with more careful control for the fact that individual observations are correlated, there is no significant difference between sites with respect to q or "a". There was some significant intersite difference for n (see below), but the findings with q and "a" indicate that the great majority of sleep time for most of our subjects was unaffected by ANEs, irrespective of site differences in nighttime aircraft noise exposure. Although most subjects seemed "tolerant" to the local ANE exposure levels during sleep, there were some whose sleep was more disturbed.

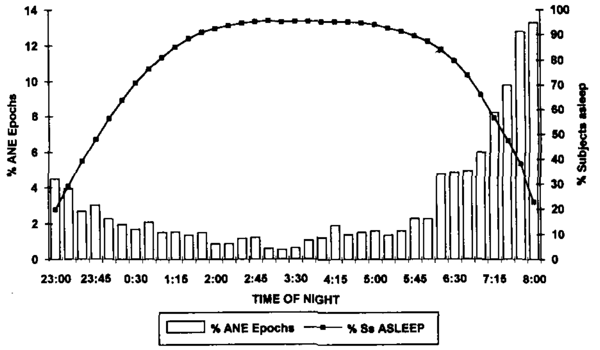


FIG. 5. Line graph: approximate percentages of subjects asleep for each 15-minute period of the night (see text for details). Histogram: percentage of sleep epochs within each 15-minute period containing aircraft noise events for all subjects.

Intersite differences in n were small [maximum intersite difference (E vs. H) = 1.52%] compared with the relatively large differences in the number of noise epochs between these sites. Nevertheless, the intersite difference in n was significant ($F = 5.33$; df = 7,390; $p < 0.001$), with the main effect being for sites C and H, which had significantly lower values than the rest.

The n values given in Table 3 are based on all ANEs occurring within the period between 2200 and 0800 hours, and between SO and reported awakening. Approximately 54% of all ANEs were below 75 dB L_{max} (outdoors), 25% were between 75 and 84 dB L_{max}, 20% were between 85 and 94 dB L_{max}, and 1% were above 95 dB L_{max}. A more elaborate breakdown of these noise characteristics is given elsewhere (16). As expected, average n − q values (all sites) rose with increasing L_{max}, which can be seen in Fig. 6. Analyses of these data (16), controlling for the confounding effects of nonacoustical factors, showed that the chance of being disturbed by ANEs below 82 dB L_{max} (outdoors) was insignificant and only became significant above this level. The time course of q over the night was similar for all sites.

Gender and age effects

There are obvious individual differences between subjects. Two such factors are age and gender. This can be seen in Fig. 7, which gives the time course of

TABLE 3. *Actimetry results for all subject nights by site, giving incidence (%) of actimeter-recorded movement onsets (actiblips) in each 30-second sleep epoch when there was aircraft noise [noise (n)] no aircraft noise [quiet (q)] and for all epochs "a"*

Site	A(N)[a]	N[a]	n[a]	A(Q)[b]	Q[b]	q[b]	A(E)[c]	E[c]	a[c]
Heathrow (A)	848	13,202	6.42	29,337	540,602	5.43	30,185	553,804	5.45
Gatwick (B)	959	15,340	6.25	34,230	613,012	5.58	35,189	628,352	5.60
Heathrow (C)	537	9,572	5.61	27,551	567,302	4.86	28,088	576,874	4.87
Gatwick (D)	1,372	22,784	6.02	34,066	640,357	5.32	35,438	663,141	5.34
Manchester (E)	2,553	35,954	7.10	33,600	611,952	5.49	36,153	647,906	5.58
Manchester (F)	1,229	18,674	6.58	31,268	547,188	5.71	32,497	565,862	5.74
Stansted (G)	267	3,913	6.82	36,639	674,265	5.43	36,906	678,178	5.44
Stansted (H)	117	2,095	5.58	32,676	647,973	5.04	32,793	650,068	5.04
All	7,882	121,534	6.49	259,367	4,842,651	5.63	267,249	4,964,185	5.38

Note that these values are for the time window 2200–0800 hours and are for the sleep period from sleep onset (algorithm) to morning awakening (sleep log).

[a] A(N) = number of actiblips in noise epochs (Lmax ≥ 60 dBA), N = number of noise epochs (Lmax ≥ 60 dBA), and n = % of noise epochs with actiblips.

[b] A(Q) = number of actiblips in quiet epochs, Q = number of quiet epochs, and q = % of quiet epochs with actiblips.

[c] A(E) = total actiblips, E = total epochs, and a = % of all epochs with actiblips.

total actiblips ("a" = n and q combined) from SO by gender and age. The small but clear differences between the sexes and, to a lesser extent, the differences between the age groups are apparent from about 180 minutes into sleep and increase towards the end of sleep. Table 4 gives the n, q and "a" values (with 95% confidence limits) by gender and age. The n − q values for men, in ascending age group order, are 1.49%, 1.92% and 0.97%. The comparable values for women are 1.0%, 0.79% and 0.77%. Analyses of variance (ANOVAs) on n and q with respect to gender are: q: $F = 15.78$; df = 1,394; $p < 0.0001$ (q higher in men) and n: $F = 12.66$; df = 1,394; $p < 0.0001$ (n higher in men). The ANOVAs for age are: q: $F = 4.73$; df = 2,394; $p < 0.009$ (q declines with age) and n: $F = 4.48$; df = 2,394; $p < 0.012$ (n declines with age).

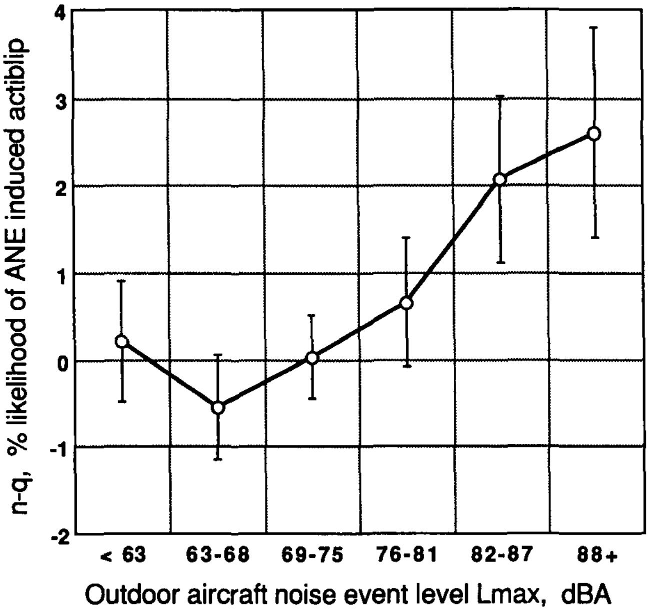


FIG. 6. Probability (n − q%), with 95% confidence intervals (based on SE) of a discrete movement (actiblip) occurring with increasing loudness (L_{max}) of the outdoor ANE. Note: Indoor levels would be about 20 dB lower.

There are no significant interactions between age and gender. It should be noted that although statistically significant, these gender and age differences are still small. For example, q values indicate that men show 10% relatively more actiblips during sleep than do women. The strength of these gender, and especially age, differences declines when more sophisticated analyses are undertaken, which control for repeated observations (16). The indication that sleep in men is more disturbed than that of women is supported by the literature (e.g. 20).

Our results show that younger people move around more in their sleep than do older people and that q diminishes with age. This is perhaps contrary to expectations, as laboratory studies suggest that sleep becomes more disturbed with age. Such studies provide single beds in comfortable rooms, but few have looked at sleep in the home, where a bed may be shared with a partner and where small children are present, as was the case for many of our younger group subjects.

End of sleep

In general terms, as well as from the EEG (e.g. 20) and earlier body movement literature (3), it may seem that as sleep progresses through the night it naturally becomes somewhat more disturbed, and perhaps "lighter". This is reflected in our data by "a" increasing over the night (Fig. 7), reaching its highest level during the last hour of sleep. However, n − q values show that sleep is not particularly liable to be disturbed by ANEs in the last hour or so of sleep. Of course, there is the possibility that a significant number of subjects

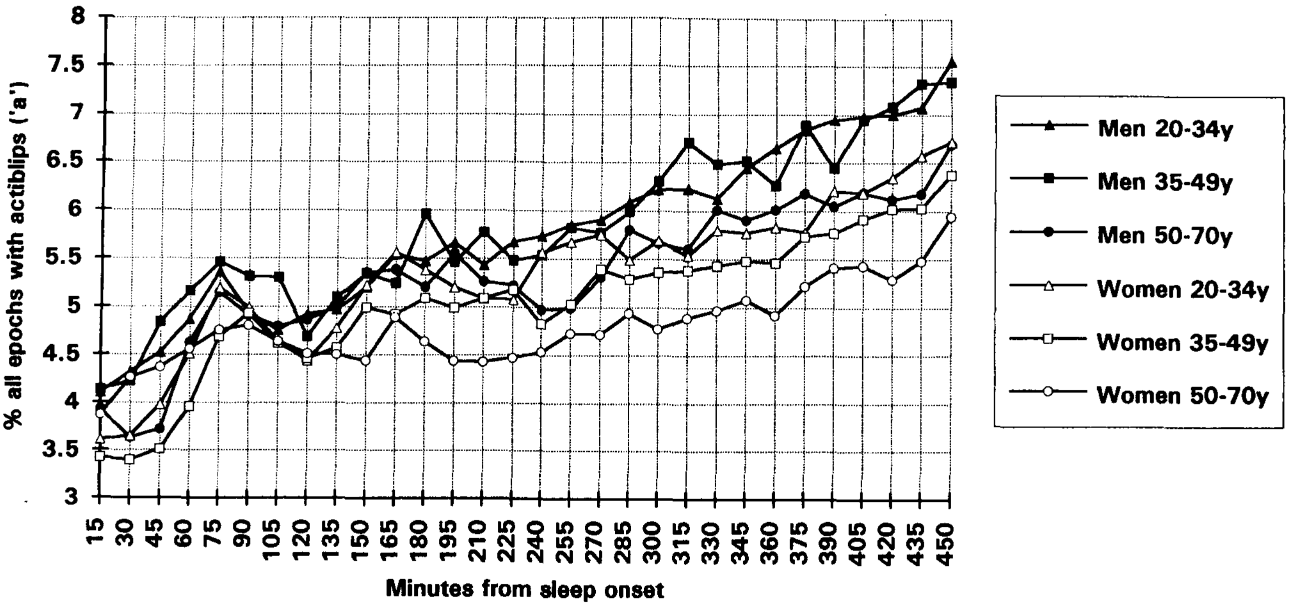


FIG. 7. Sleep trends for "a" (percentage of epochs containing an actiblip, n and q combined) in 15-minute periods, by gender and age. Men have significantly higher values than do women, and there is a trend for a decline with age.

are (prematurely) woken up by an ANE towards the end of their sleep, who then get up and remove themselves from the sample, leaving only unresponsive sleepers in the pool. Another possibility is that having been awoken, subjects lie awake, move around more, but do not respond to a specific ANE (hence n would fall and q would rise, giving a low $n - q$). The relatively few subjective reports of awakenings by ANEs indicate that premature morning awakenings are uncommon.

Subjective responses in the morning sleep logs

Figure 8 gives the correspondence between q (all nights) and the associated subjective sleep quality reports from the following mornings' logs. The total number of nights when the actimeter was worn and the subject remembered to complete the log was 5,525. Numbers above each bar give the subject nights in that sample. The very small standard error bars are also shown. The association between these two variables is high and is significant at $p < 0.01$ ($F = 13.0$; df = 4,5517). Such findings further support the validity of using actimeters. When the same question was given to the same subjects, using the same scale values as those for the morning log, but in a one-off version as part of the initial social survey questionnaire administered before the main study, the association between the categories of reply and the values for q was not significant; there was no clear increment of q over the values 1 to 5 for the questionnaire, as is the case in Fig. 8. As one might expect, by asking subjects every day about sleep quality, via a sleep log, one can obtain more reliable data about their sleep (using actimetry as the benchmark) than by asking the same question once in a survey.

In all, out of the total of 5,716 subject nights when

TABLE 4. *Gender and age incidences for actiblips in sleep. Mean n, q and a values with 95% confidence limits for percent of epochs showing an actiblip response after aircraft noise (n), with no aircraft noise (q) and for all epochs (a = n and q combined)*

Sex	Age group (years)	n[a]	Lower	Upper	q[b]	Lower	Upper	a[c]	Lower	Upper
Male	20–34	7.24	6.89	7.58	5.75	5.70	5.80	5.79	5.74	5.84
Male	35–49	7.67	7.26	8.09	5.75	5.70	5.81	5.80	5.74	5.85
Male	50–70	6.29	5.93	6.64	5.32	5.26	5.37	5.34	5.29	5.39
Female	20–34	6.41	6.11	6.70	5.41	5.36	5.45	5.43	5.39	5.48
Female	35–49	5.87	5.57	6.18	5.08	5.04	5.13	5.10	5.06	5.15
Female	50–70	5.62	5.28	5.95	4.85	4.80	4.90	4.87	4.82	4.92
All		6.49	6.35	6.62	5.36	5.34	5.38	5.38	5.36	5.40

[a] n = % of noise epochs with actiblips.
[b] q = % of quiet epochs with actiblips.
[c] a = % of all epochs with actiblips.

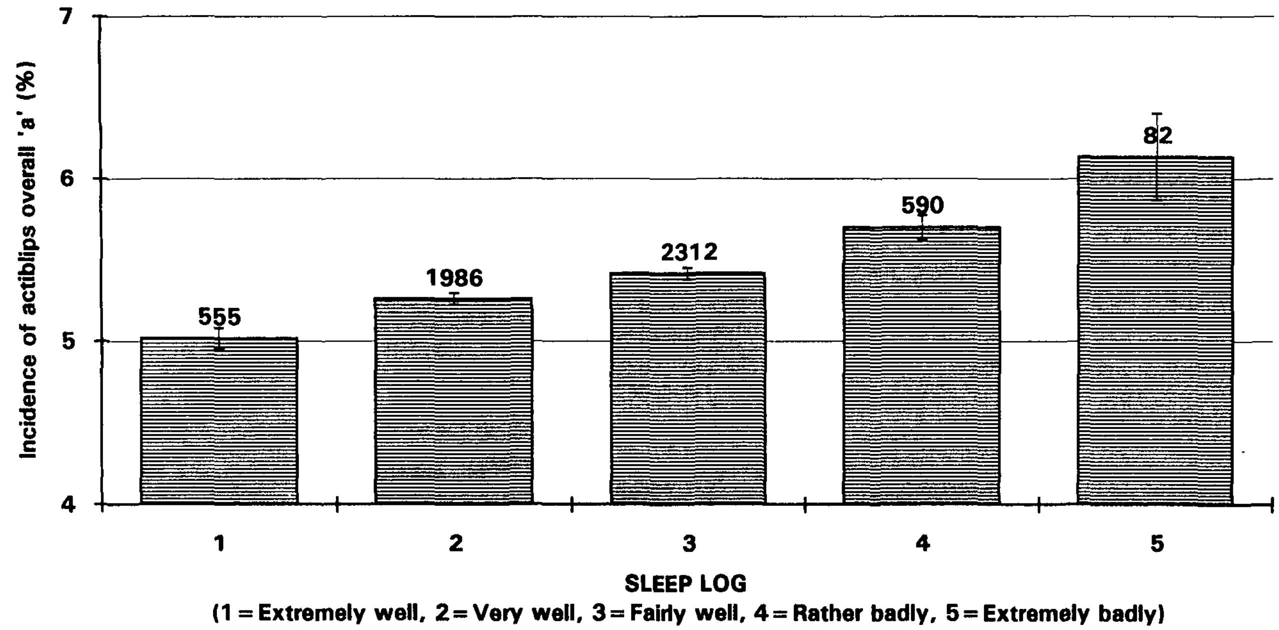


FIG. 8. Mean incidence of all actiblips ("a") for each subject night shown against each morning sleep log response to the question: "Last night I slept . . ." (5,525 nights). Standard error bars and number of nights involved are shown for each value. The correlation is significant; the greater the movement, the poorer the subjectively reported sleep.

the log was completed (95.3% response), 3,546 contained reports of at least one awakening, with 6,457 reported awakenings in all. Figure 9 gives the relative numbers of awakenings, ranked for the main categories of awakening. The most frequent overall response was "don't know", with "toilet" and "children" being the next most likely. "Toilet" was the predominant response in both men and women from oldest group, and "children" was the predominant response in the group of youngest women. ANEs were a comparatively minor cause and were least likely to affect the youngest group. Out of the 400 subjects, 97 (24.2%) reported being awoken by an ANE on at least one occasion on at least one night. In total this came to 284 subject nights (all sites), which is about 5% of all nights. Of these 284, ninety-five came from the noisiest site (E), resulting in about a 13% likelihood of a subject reporting being awoken on any night by an ANE at this site. Such likelihoods for the remaining seven sites, in rank order of nighttime ANE noise levels (cf. Table 3), are: 3% (site D), 6% (site F), 3% (site B), 5% (site A), 1% (site C), 3% (site G) and 4% (site H).

The total number of awakenings attributable to ANEs was 351 (=5.4% of all 6,457 subjective awakenings, i.e. some subjects reported more than one awakening for a night). The "don't know" category must contain some less subjectively obvious ANEs, as well as other actual causes. Whether or not "don't know" contains a disproportionately high number of hidden ANE awakenings is a matter that cannot be resolved. But from the n − q data and the findings given below, we have little evidence of this.

Given that our overall n − q was 1.13% and that the total number of ANEs that could have been heard (i.e. N) over all subject nights (from 2200 to 0800 hours) was 121,534 (Table 3), this gives about 1,373 actual ANEs responded to actimetrically—almost quadruple those reported subjectively. However, only a portion of these actiblips would have been an actual awakening. Our EEG–actigram comparisons showed that on an average night only 40% of actiblips indicated actual awakenings. Forty percent of the estimated actiblip responses to ANEs is 549 putative awakenings, which can be compared with the subjectively reported 351 responses.

CONCLUSIONS

The subjects in this field study were less responsive to noise during sleep than has been reported from laboratory studies. This is perhaps because of the artificial nature of laboratory studies in which subjects are largely unadapted to the physical surroundings (Fig. 1). Our subjects were typical of the adult populations of the locations investigated, and it must be assumed that because they had been living where they were for over 1 month (usually for several years), they must have acquired some degree of adaptation to ANEs during their sleep. Few of our initial survey respondents who

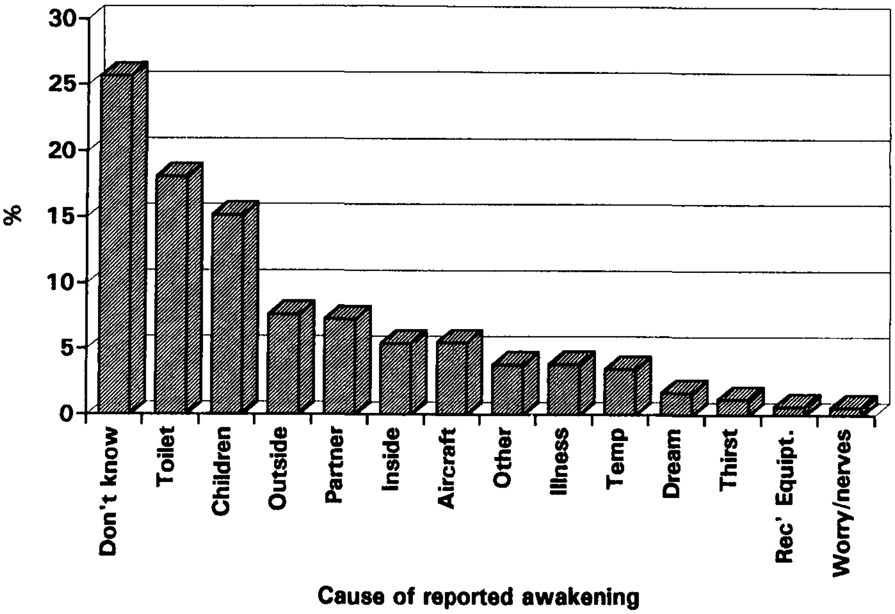


FIG. 9. Causes and prevalence of all awakenings reported in morning sleep logs for all nights, by gender and age. Total awakenings = 6,457. Each subject could have reported more than one awakening per night.

volunteered for the main study were specifically excluded. Because our subjects lived near airports, they may not be typical of the wider adult population. Ideally, additional control sites away from airports would have been included. However, because of the extreme difficulty of selecting sites that adequately matched the relevant but still largely indeterminate characteristics (other than aircraft noise) of the test sites and because the quietest sites (at G and H) experienced little nighttime ANEs on average (with nil ANEs on some nights), it was considered at the outset that the inclusion of control sites would not make the best use of our constrained budget.

The following main conclusions can be drawn:

1. Actimetry is a valid and sensitive measure of awakenings during sleep, especially those lasting for >15 seconds. The EEG recordings displayed an average of 16 of such events per night, with 88% of these being detected by actimeters. However, there were about 2.5 times this number of actiblips per night, indicating that more minor arousals were also being detected, as well as the lesser limb movements of sleep that have no bearing on sleep quality. Our finding of an average of about 47 individual movements during sleep (all subjects) falls mid-range of the 20–60 movements per night reported by Kleitman (3), who found that more than half of these movements were minor.

2. By actimetry and self-report, the sleep of most subjects was largely unaffected by ANEs. Of course, the louder the ANE, the greater the likelihood of an effect on sleep; but nevertheless, the responses to louder ANEs (e.g. L_{max} >80 dB, outdoors) were still very low on average. In fact, for the great majority of our subjects the most disturbing influences on sleep were not ANEs but more idiosyncratic factors such as young children, illness, a need to go to the toilet and the bed partner. Assuming an average outdoor–indoor sound level difference of about 20 dB, these findings are largely in agreement with those of Griefahn (21), who has shown in her field studies that the probability is near to zero of being awoken by noise less than 60 dB as measured indoors. We found (Fig. 6) that ANEs of about 82 dB L_{max} or less (outdoors) caused little, if any, sleep disturbance. The general tolerance of many people to ANEs during sleep was also indicated in an initial pilot study we undertook at the airport with the most night movements (Manchester), which fortuitously closed for runway repairs at night for 10 nights during the study. This had little effect on the EEGs and actimetrically measured sleep of 16 people (age range 30–70 years), who habitually lived very close to the main runway. Most of them were quite unaware of this closure before, during and after the study. A fuller report of this finding is in preparation.

3. There were significant individual differences in response to ANEs, particularly with regard to gender and, to some extent, age.

4. Despite large site differences in the number of nighttime ANEs, the difference between sites in ANE effects on actimetrically determined limb movements was very small.

5. There was little subjective or objective indication to show that our oldest subject group (50–70 years) had poorer sleep than the youngest group (20–35 years),

although the former tended to report more awakenings by ANEs. Actimetry showed the latter to have more discrete movements (q) during sleep.

Acknowledgements: We thank the following people for their help with data collection and analysis: Brian Atkinson, Roy Cadoux, Ben Emm, Jon Forster, Ceril Jones, Ian Jopson, Heather Lomas, John Ollerhead, Fiona Van, Alison Watson and Alan Woodley. This project was financed by the U.K. Department of Transport. Certain of the tables and figures originally appeared in an earlier, unpublished report (16).

APPENDIX

Sleep onset (SO) from actigrams

To gauge SO from the actigrams, the 178 concurrent EEGs were used together with a further data base of 53 nights of EEG and synchronized actimeter data obtained from two earlier pilot studies. These latter data largely comprised subjects in the oldest age group (50–70 years). Most subjects contributed 3 or 4 nights of data. By age range these nights were broken down as: 20–34 years = 55; 35–49 years = 62 and 50–70 years = 114. Our EEG criterion for SO was, "following lights out, the onset of the first 6-minute period or longer of continuous sleep consisting of stage 2 (18) or deeper". Having located this on each hypnogram, the same time point was then identified on the corresponding actigram. An actigram algorithm was then derived from the EEG-based SO. The one giving the best fit was based on the onset of a period of at least 14 actigram epochs (7 minutes) of nil movement after "lights out". On average, the EEG-determined SO criterion was met after 10 epochs (5 minutes) into this 7-minute period. This "14,10" algorithm gave a mean hit (for all 231 nights) on the SO criterion of −4 seconds (SD = 13.4 minutes). Figures 2 and 3 illustrate the SO tagging on the raw and filtered actigrams.

Despite further analyses on an age group by gender basis, we found no reason to prescribe different SO algorithms on this basis (the mean and SD for the three age groups in ascending order of age were 3 seconds, 14.9 minutes; −1 second, 12.3 minutes and 2 seconds, 13.0 minutes); that is, individual differences were more idiosyncratic than age or gender criteria. ACCORD then tagged all 5,742 actigrams at the estimated SO point.

REFERENCES

1. Pearsons KS, Barber DS, Tabachnick BG. *Analyses of the predictability of noise-induced sleep disturbance.* Report no. HSD-TR-89-029. Ohio: Wright-Patterson Air Force Base, 1990.
2. Muzet A, Naitoh P, Townsend RE, Johnson LC. Body movements during sleep as a predictor of state change. *Psychon Sci* 1972;29:7–10.
3. Kleitman N. *Sleep and wakefulness.* Chicago: University of Chicago Press, 1963.
4. Tryon WW. *Activity measurement in psychology and medicine.* New York: Plenum Press, 1991.
5. Middelkoop HAM, van Hilten BJ, Kramer CGS, Kamphuisen HAC. Actigraphically recorded motor activity and immobility across sleep cycles and stages in healthy male subjects. *J Sleep Res* 1993;2:28–33.
6. Griefahn B. Sleeping in noisy environments: effect, limits, and preventative measures. In: JA Horne, ed. *Sleep '90.* Bochum: Pontenagel Press, 1990:391–3.
7. Libert JP, Bach V, Johnson LC, Ehrhart J, Wittersheim G, Keller D. Relative and combined effects of heat and noise exposure on sleep in humans. *Sleep* 1991;14:24–31.
8. Mullaney DJ, Kripke FF, Messin F. Wrist-actigraphic estimation of sleep time. *Sleep* 1980;3:83–92.
9. Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An activity-based sleep monitor system for ambulatory use. *Sleep* 1982;5:389–99.
10. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake detection from wrist activity. *Sleep* 1992; 15:461–9.
11. Dunham DW, Hoffmann RF, Broughton RJ. Wrist activity and sleep/wake estimation revisited. *Sleep Res* 1991;20A:491.
12. Sadeh A, Alster J, Urbach D, Lavie P. Actigraphically bedtime sleep–wake scoring: validity and clinical applications. *J Amb Monitor* 1989;2:209–16.
13. Cole RJ, Kripke DF. Progress in automatic sleep/wake scoring by wrist actigraph. *Sleep Res* 1988;17:331.
14. Hauri PJ, Wisbey J. Wrist actigraphy in insomnia. *Sleep* 1992; 15:293–301.
15. Brooks JO, Friedman L, Bliwise DL, Yesavage JA. Use of the wrist actigraph to study insomnia in older subjects. *Sleep* 1993; 16:151–5.
16. Ollerhead JB, et al. Report of a field study of aircraft noise and sleep disturbance. U.K. Department of Transport, 1992.
17. Kryger MH, Roth T, Dement WC. *Principles and practice of sleep medicine.* Philadelphia: W. B. Saunders, 1989, 739 pp.
18. Rechtschaffen A, Kales A. *A manual of standardized terminology, techniques, and scoring system of sleep stages in human subjects.* Los Angeles: UCLA Brain Information Service, 1968.
19. Diamond I, Jones C, MacKean J, Ollerhead J. Sleep disturbance due to aircraft noise: social survey report. Report SRSG(92) 8. Southampton University Department of Social Statistics, 1992.
20. Williams RL, Karacan I, Hursch CJ. *Electroencephalography (EEG) of human sleep: clinical applications.* New York: Wiley, 1974.
21. Griefahn B. Noise induced sleep disturbance and the effects on health. *J Sound Vib* 1978;59:99–106.