Available online at www.sciencedirect.com

ScienceDirect

Journal of Sound and Vibration 302 (2007) 197–207

JOURNAL OF SOUND AND VIBRATION

www.elsevier.com/locate/jsvi

A-weighted sound pressure level as an indicator of short-term loudness or annoyance of road-traffic sound

Mats E. Nilsson[a,b,*]

[a]*Institute of Environmental Medicine, Karolinska Institutet, Sweden*
[b]*Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden*

Received 5 February 2006; received in revised form 17 November 2006; accepted 20 November 2006
Available online 11 January 2007

Abstract

Two listening experiments were conducted in order to determine whether *A*-weighting is a valid indicator of the perceived loudness or annoyance of road-traffic sound. Because *A*-weighting has been criticized for not properly integrating energy at low frequencies, experimental road-traffic sounds were selected with a wide range in low-frequency content, assessed as the difference between *C*- and *A*-weighted sound pressure levels (L_{C-A}). In the first experiment, 30 listeners assessed the perceived loudness of the selected sounds. In the second experiment, another group of 31 listeners assessed the perceived annoyance of the same sounds. Sounds with high levels of L_{C-A} were louder and more annoying than sounds with medium levels of L_{C-A}, which in turn were louder and more annoying than sounds with low levels of L_{C-A}, at similar *A*-weighted sound pressure levels (L_A). It was estimated that the change in perceived loudness or annoyance associated with a 1 dB change in L_{C-A} would correspond to approximately a 0.4 dB change in L_A. In contrast, sounds with similar Zwicker loudness levels (L_N) were approximately equal in loudness and annoyance irrespective of their L_{C-A}. Thus, L_N was found to be superior to L_A as an indicator of short-term loudness and annoyance of road-traffic sounds with wide variation in low-frequency content.

1. Introduction

In most national standards, road-traffic sound is assessed in terms of its *A*-weighted sound pressure level. A common critique of *A*-weighting is that it overcompensates for the hearing system's reduced sensitivity at low frequencies [1–3]. Road-traffic sounds may vary considerably in low-frequency content due to variations in source characteristics and distance between receiver and vehicles. It may therefore be questioned if *A*-weighting is a valid indicator of the perceived loudness or annoyance of road-traffic sound. This question was addressed in two experiments, in which listeners assessed short-term loudness (Experiment 1) or annoyance (Experiment 2) of road-traffic sounds with a wide variation in low-frequency content.

Most previous experimental studies on the perception of road-traffic sound have failed to show that *A*-weighting is inferior to alternative indicators of perceived loudness or annoyance [4–9]. One probable

---

[*] Corresponding author. Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden. Tel.: +46 8 6121404.
E-mail address: mats.nilsson@psychology.su.se.

0022-460X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2006.11.010

reason for the relative success of A-weighting in these experiments is that they were not designed to truly put A-weighting to the test. A critical test would include several sounds with the same A-weighted sound pressure level but with a wide range of low-frequency content. To my knowledge, Watts ([10], third experiment) is the only previous study on road-traffic sound that conducted such a test. He accomplished this by filtering recordings of road-traffic sound from single vehicles. The perceived noisiness of sounds with similar A-weighted sound pressure levels was found to differ consistently with the low-frequency content. Furthermore, the loudness level [11,12], was found to be superior to A-weighted sound pressure level as an indictor of perceived noisiness.

The present experiments expand Watts's [10] study in two respects. First, Watts did not quantify the effect of low-frequency content on perceived noisiness. However, quantification is needed in order to estimate the size of the prediction error attributed to the use of A-weighting. In the present study, the difference between C- and A-weighted sound pressure level (L_{C-A}) was used as the measure of low-frequency content (e.g., Refs. [13,14]). Following Vos [15], the effect of low-frequency content could then be quantified by including L_{C-A} in a multiple regression model.

Second, Watts [10] manipulated recordings of road-traffic sound. The $\frac{1}{3}$-octave-band sound levels between 1 and 2.5 kHz were increased up to 15 dB, while keeping the overall A-weighted sound pressure level (ASEL) constant. Such sounds are unlikely to be heard in real life. The present experiments were designed to include non-manipulated sounds taken from real recordings. At the same time, the selected sounds also had a wide variation in low-frequency content. In order to achieve this, a "stimulus population" was created by dividing several hours of recordings of road-traffic sound into 3-s excerpts. Each excerpt was measured with respect to overall A-weighted sound pressure level and L_{C-A}. Experimental sounds were then selected in order to obtain a set of road-traffic sounds varying in L_{C-A} at equal A-weighted sound pressure levels.

2. Experiment 1: loudness

2.1. Method

2.1.1. Binaural recordings

Binaural recordings of approximately 45 min were conducted in 12 gardens of a residential area of two-story houses, located along a main road (19,600 vehicles/24 h on weekdays; 6% heavy vehicles). The recordings were conducted in 1999 always during the afternoon rush hours. The traffic was not congested; the average speed was 60 km/h. Due to stop signs, the traffic flow was irregular and contained periods with vehicles at regular speed, as well as, periods with accelerating and decelerating vehicles. The road was flat and its surface was conventional asphalt. There were no screens between microphone and road. The distance between the center of the road and the microphones was 23 m. The microphone height was 1.6 m.

2.1.2. Selection of experimental sounds from a stimulus population

The 45-min binaural recordings were edited in order to eliminate periods with major wind disturbances, sounds from people talking close to the microphones, etc. After the editing, the recordings had a total duration of 6 h and 25 min. These recordings were divided into 3-s excerpts using a program written in MATLAB. In order to obtain a very large set of excerpts, the 3-s excerpts were sampled each second of the recording (i.e., there was a two second overlap between adjacent excerpts). This procedure generated a "stimulus population" of 23,114 excerpts. The duration of the excerpts corresponded to the duration of the "psychological present" [16,17]. This assured that assessments referred to short-term perceptions of the traffic sounds and therefore were unaffected by various cognitive biases known to affect retrospective assessments of longer periods [18].

A-weighted sound pressure level ($L_{A\mathrm{eq},3\,\mathrm{s}}$, in the following denoted L_A) and the difference between A- and C-weighted sound pressure levels ($L_{C\mathrm{eq},3\,\mathrm{s}}$–$L_{A\mathrm{eq},3\,\mathrm{s}}$, in the following denoted L_{C-A}) was determined for each excerpt. These measurements referred to the channel of the binaural recoding with the higher value. The Pearson coefficient of correlation between L_A and L_{C-A} in the stimulus population was -0.49.

The selection of experimental sounds from the 23,114 excerpts was conducted in the following way:

1. All excerpts in the stimulus population were categorized in 38 groups of equal L_A (± 0.5 dB), from 43 to 85 dB.
2. The experimental sounds were selected from 16 of these L_A-categories, from 48 to 78 dB in 2-dB steps.
3. From each of these 16 categories, three experimental sounds were chosen representing sounds with low, medium and high levels of L_{C-A}. This was accomplished by random selection among the 10% of excerpts with lowest values of L_{C-A}, among the excerpt(s) with the median value of L_{C-A} and among the 10% of excerpts with highest values of L_{C-A}, respectively.

Thus, $16 \times 3 = 48$ experimental sounds were chosen. Although only 3-s long, the sounds were perceived as realistic sounds dominated by traffic noise from a nearby major road. The selected sounds were all from different periods in the recordings; there was no overlap in time between any two of the selected sounds.

2.1.3. *Acoustic analysis of experimental sounds*

The selection of experimental sounds described above was based on sound levels referring to the channel, left or right, with the greater value of L_A. Analysis of the absolute binaural sound level difference, $\Delta L = |L_{A,\text{left-channel}} - L_{A,\text{right-channel}}|$, revealed systematic differences between the three groups of excerpt sounds. On average, ΔL was 1.8, 1.1, and 0.7 dB for excerpts with small, medium and high relative levels of low-frequency content, respectively. In order to account for this difference, an average was calculated on the antilog of decibel-values from each channel, and the logarithm of this average was used in all analyses reported below. Fig. 1 shows the resulting sound levels (L_A) and relative levels of low-frequency content (L_{C-A}) for the 48 chosen road-traffic sounds. For sounds with approximately equal levels of L_A, the maximum range in L_{C-A} was 15 dB (vertical distance between filled circles and triangles in Fig. 1).

Along with L_A, L_{C-A} and ΔL, the experimental sounds were also analyzed with respect to Zwicker loudness level (ISO 532B [11]) calculated from the $\frac{1}{3}$-octave spectra averaged over the 3-s duration of the sound (L_N), A-weighted sound pressure level and loudness level exceeded 10% of the time (L_{A10} and L_{N10}) and the difference between levels exceeded 10% and 90% of the time (L_{A10-90} and L_{N10-90}). All these measures refer to the logarithmically averaged level across the left and right channel of the binaural recording. Ranges and inter-correlations of the acoustical variables are given in Table 1.

In addition to the 48 experimental traffic sounds, 21 sound levels of pink noise, ranging from 44 to 84 dBA in 2 dB steps, were used as reference sounds in accordance with Berglund's master-scaling method [19]. The main idea of this method is to provide a well-defined context of reference sounds, within which a set of "target" sounds may be assessed. This allows for comparison between studies with the same set of reference sounds but different target sounds. The use of well-defined reference sounds also makes it possible to express

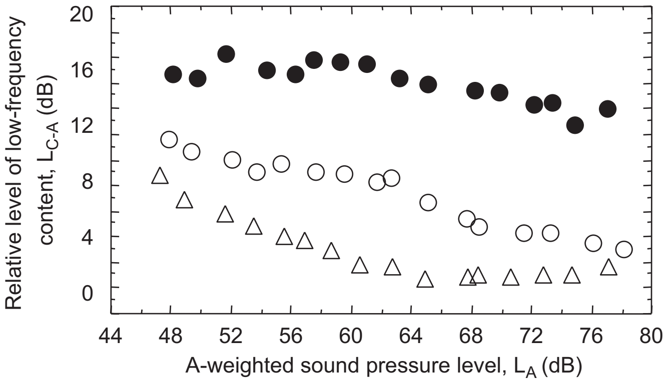

Fig. 1. A-weighted sound pressure level (L_A) and low-frequency content (L_{C-A}) of the experimental road-traffic sounds. ● High, ○ medium and, △ low relative level of low-frequency content.

Table 1
Ranges and inter-correlations of acoustical variables

	Pearson coefficient of correlation (r)							
Range (dB)	Variable	$L_{C-\mathrm{A}}$	L_A	L_{A10}	L_N	L_{N10}	L_{A10}–L_{A90}	L_{N10}–L_{N90}
1–18	L_{C-A}							
47–78	L_A	−0.32						
48–80	L_{A10}	−0.31	>0.99					
65–96	L_N	−0.17	0.99	0.99				
67–97	L_{N10}	−0.15	0.98	0.98	>0.99			
1–12	L_{A10}–L_{A90}	−0.13	0.33	0.40	0.34	0.40		
1–10	L_{N10}–L_{N90}	0.04	0.10	0.17	0.13	0.21	0.89	
0–4	ΔL	−0.47	0.23	0.25	0.17	0.17	0.28	0.16

magnitude estimates in a meaningful unit, namely, the physical unit of the reference sound (see below, Section 2.2.1). In the present application, the reference sounds were pink noises, and the targets were traffic sounds.

2.1.4. *Listeners*

Thirty university students participated in the experiment (25 women, 5 men; mean age = 29 years). All participants' hearing threshold levels [20] were below 25 dB in their best ear in all tested frequencies (0.5, 1, 2, 3, 4, 6, and 8 kHz, pure-tone audiometer: Brüel & Kjær, Type 1800). The listeners received course credit for their participation.

2.1.5. *Procedure*

The experiment consisted of eight listening sessions. In each of these sessions, 69 sounds were presented, including the 21 pink-noise references and the 48 experimental traffic sounds (Fig. 1). Experimental and reference sounds were presented in random orders, which were different for each session and participant. Listening session duration was approximately 7 min. The sessions were separated from one another with 2-min pauses, except for the pause between sessions 4 and 5, which was 10 min. A training session with 12 sounds was conducted prior to the first session. Perceived loudness was assessed with the method of free number magnitude estimation [21]. The participants entered their magnitude estimates on a computer keyboard.

2.1.6. *Equipment*

The binaural recordings were conducted using a DAT-recorder (Sony TCD-D10 ProII, A/D: 16 bit, 48 kHZ) and a binaural recording system (Artificial head Brüel & Kjær Type 4100, with Brüel & Kjær microphones Type 4190, preamplifiers Type 2669 and a microphone amplifier Type 2690 NEXUS). The recordings were edited on a personal computer (Dell Precision 220). Unusable recording periods (e.g., due to wind disturbances) were first manually removed from the original recordings using a sound-editing software (Sound Forge 7.0). A large number of 3-s excerpt sounds were then created and analyzed with respect to average $\frac{1}{3}$-octave-band levels using scripts written in MATLAB.

The listener was seated in a soundproof room in front of a computer screen connected to a computer in another room. In order to guarantee good reproduction of low frequencies, the excerpt sounds were simultaneously reproduced on a pair of earphones (Sennheiser HD 600) and a pair of subwoofers (Velodyne). Measurement of our earphones showed a drop in the frequency response below 40 Hz. Therefore, the earphones were set to reproduced frequencies above 40 Hz, whereas the subwoofers reproduced frequencies in the range of 15–40 Hz. The whole listening system was calibrated using a pink-noise signal, which was measured at the point of the listener's ear. The digital signal, stored on a personal computer (Dell Precision 220; Sound card: LynxTwo), was fed into a digital filter and D/A-converter with a four-channel analogue output (Rane RPM 26z). One set of digital filters was applied to the stereo signal reproduced by the earphones. These filters included a high-pass filter at 40 Hz as well as filters correcting for the nonlinear frequency response of the earphones in the higher frequencies. A digital low-pass filter at 40 Hz was applied to the stereo signal reproduced by the subwoofers. The subwoofers were also equipped with an analogue low-pass filter, which was set to 40 Hz. The frequency response of the whole listening system was flat within 2 dB ($\frac{1}{3}$-octave-band levels, 20–16 000 Hz).

The experimental sounds were randomly ordered and presented using a MATLAB script. The same script was used for collecting the listener's responses.

2.2. Results and discussion

The consistency of individual data was found to be high, as determined by the coefficients of correlation (Pearson) between magnitude estimates of sounds in the first and last four sessions. All except two participants had coefficients greater than 0.85 for the pink-noise reference-sounds (median = 0.97, range: 0.62–0.99) and all except two participants had coefficients greater than 0.85 for the traffic sounds (median = 0.91, range: 0.57–0.96). The general trend of the data of the participants with coefficients below 0.85 was not different from the other participants. Therefore, data from all participants were included in the group analyses.

2.2.1. Pink noise equivalent sound level (PNE$_{loud}$)

The perceived loudness of each traffic sound was expressed as the pink-noise equivalent sound level (PNE$_{\text{loud}}$). The PNE$_{\text{loud}}$ of a traffic sound is the sound level of an *equally loud* pink noise. The main advantage of expressing loudness as PNE$_{\text{loud}}$ is (a) that it gives loudness a meaningful unit (pink-noise sound level in dBA), and (b) that it does not presuppose that listeners are able to produce magnitude estimates with ratio-scale properties. The only assumption is that, on average, equal numbers (magnitude estimates) means equal loudness (cf. Refs. [19,22,23]).

PNE$_{\text{loud}}$-values were determined by first calculating the geometric mean magnitude estimate (R_{pn}) for each listener and pink-noise sound level (L_{pn}). These geometric means were used to derive individual psychophysical functions,

$$\log(R_{\text{pn}}) = a + bL_{\text{pn}}, \tag{1}$$

where a and b are constants unique to each listener. The fits, R^2, of these individual functions were excellent, between 0.86 and 0.99 (median = 0.97). The slope, b, of the functions varied between 0.006 and 0.055, median = 0.018.

Second, for each listener and road-traffic sound, the geometric mean magnitude estimate (R_{traffic}) was calculated. These geometric means were then transformed into pink-noise equivalent sound levels, using each listener's unique set of constants (a and b). The logic behind the transformation was as follows: Equal loudness of a traffic sound and a pink-noise sound level would imply that a listener, on average, would give the two sounds the same magnitude estimate. Thus, $R_{\text{pn}} = R_{\text{traffic}}$, and, from Eq. (1),

$$\log(R_{\text{traffic}}) = a + bL_{\text{pn}}. \tag{2}$$

The PNE$_{\text{loud}}$ equals the sound level of pink noise in Eq. (2). Thus, $\text{PNE}_{\text{loud}} = L_{\text{pn}}$, and, from Eq. (2),

$$\text{PNE}_{\text{loud}} = [\log(R_{\text{traffic}}) - a]/b. \tag{3}$$

Eq. (3) was used to calculate individual PNE$_{\text{loud}}$-values for each traffic sound. The standard deviation of individual PNE$_{\text{loud}}$-values for a given traffic sound ranged between 3.2 and 9.3 dB PNE$_{\text{loud}}$ (mean = 5.4 dB, $N = 48$ traffic sounds). The underlying distributions were approximately symmetric with the difference between the arithmetic mean and the median PNE$_{\text{loud}}$-values ranging between −2.1 and 0.9 dB PNE$_{\text{loud}}$ (mean = −0.6 dB, $N = 48$ traffic sounds). Therefore, it was justified to use arithmetic means for summarizing the results across listeners.

2.2.2. Psychoacoustic relationships for perceived loudness (PNE$_{loud}$)

If L_A was a valid indicator of perceived loudness, then equal levels of L_A would mean approximately equal levels of perceived loudness. However, this was not the case, as seen in Fig. 2(a), where PNE$_{\text{loud}}$-values averaged across listeners are shown as a function of L_A. The line shows the linear regression function; the corresponding equation, fit and standard error of the estimates are given in the second row of Table 2. In general, sounds with high levels of L_{C-A} (filled circles) were louder than sounds with medium levels of L_{C-A} (open circles), which in turn were louder than sounds with low levels of L_{C-A} (triangles). The absolute values

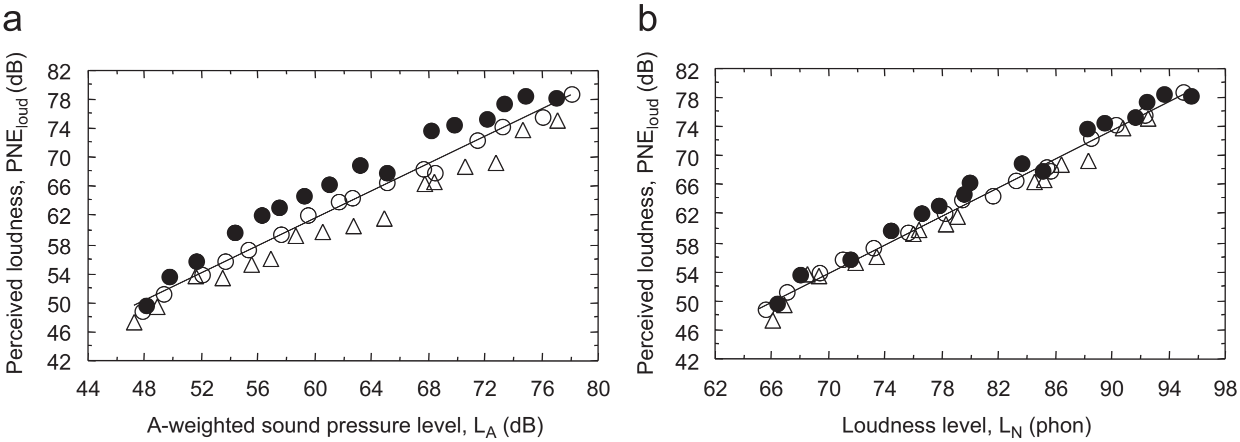

Fig. 2. Perceived loudness (PNE_{loud}) as a function of: (a) A-weighted sound pressure level, L_A, and (b) loudness level, L_N (ISO 532B [11]). ● High, ○ medium and, △ low relative level of low-frequency content.

Table 2
Linear relationships between perceived loudness (PNE_{loud}) and acoustical variables

Empirical relationship	R^2	Standard error (dB PNE_{loud})
$PNE_{loud} = 5.38 + 0.94L_A$	0.92	2.6
$PNE_{loud} = -3.80 + 1.03L_A + 0.42L_{C-A}$	0.99	1.1
$PNE_{loud} = -15.30 + 0.99L_N$	0.98	1.4

of PNE_{loud} were higher than the L_A-values of road-traffic sounds with high levels of L_{C-A} (filled circles), and roughly equal to the L_A-values of traffic sounds with medium or low levels of L_{C-A} (open symbols). This means that, for similar L_A-values, pink noise was less loud than traffic sounds with high levels of L_{C-A}, but roughly equal in loudness to traffic sounds with low or moderate levels of L_{C-A}. This makes sense, since pink-noise has a fairly low level of L_{C-A} (approx. 2 dB).

In order to determine the effect of low-frequency content on perceived loudness, a standard multiple regression analysis was conducted with PNE_{loud} as the dependent variable and L_A and L_{C-A} as the independent variables. The empirical equation obtained is given in the third row of Table 2, together with the fit and standard error of the estimate. All coefficients were significantly greater than zero ($p < 0.001$). L_{C-A} was found to explain a large part of the variance in perceived loudness not accounted for by L_A. The model fit, R^2, increased from 0.92 with only L_A in the model to 0.99 with both L_A and L_{C-A} in the model. The ratio of the coefficients for L_{C-A} to L_A was 0.41 (0.42/1.03), suggesting that a 1 dB change in L_{C-A} has the same effect on loudness as a change in L_A of approximately 0.4 dB.[1]

---

[1] A sequential regression analysis was employed to determine if the effect of L_{C-A} on perceived loudness (Experiment 1) or annoyance (Experiment 2) was in fact caused by a variation in the binaural level difference (ΔL) acting as a "confounding" variable (cf. Ref. [28]). This could be suspected because ΔL and L_{C-A} were correlated [$r = -0.47$ (Table 1)]. In the first step of the sequential regression analysis, ΔL was entered as the only independent variable in the regression model. In the second step, L_A and L_{C-A} were added to the regression model, accounting for the variance not accounted for by ΔL. In Experiment 1, the regression coefficient for ΔL (-0.28) was found to be insignificant ($p = 0.136$), whereas the coefficients for L_A (1.03) and L_{C-A} (0.40) were significant ($p < 0.001$). The coefficient for L_{C-A} was only slightly lower than the coefficient obtained without controlling for ΔL. The ratio of the coefficients of L_{C-A} and L_A was 0.39, which is similar to the ratio of 0.41 obtained without controlling for ΔL. In Experiment 2, the regression coefficient for ΔL (-0.01) was found to be insignificant ($p > 0.05$), whereas the coefficients for L_A (1.27) and L_{C-A} (0.45) were significant ($p < 0.001$). The coefficient for L_{C-A} was only slightly lower than the coefficient obtained without controlling for ΔL. The ratio of the coefficients of L_{C-A} and L_A was 0.36, which is the same as obtained without controlling for ΔL. It may be concluded that the observed effect of L_{C-A} on perceived loudness (Experiment 1) and annoyance (Experiment 2) was not the result of a variation in ΔL. Observe that the sound levels, L_A and L_{C-A}, used in the analysis were combined levels of the left and right channels of the binaural recordings (see above Section 2.1.3). The result of the sequential regression analysis shows that this successfully compensated for the difference in ΔL between the experimental sounds.

Fig. 2(b) shows $\mathrm{PNE}_{\mathrm{loud}}$ as a function of loudness level, L_N. The line shows the linear regression function; the corresponding equation, fit and standard error of the estimates are given in the fourth row of Table 2. The fit and standard error show that L_N was superior to L_A as an indicator of perceived loudness of the traffic sounds. In fact, the standard error of the estimates was only slightly higher for L_N than for the model including both L_A and L_{C-A}, 1.4 and 1.1 dB $\mathrm{PNE}_{\mathrm{loud}}$, respectively (see Table 2). Observe that $\mathrm{PNE}_{\mathrm{loud}}$ is expressed in dBA (see Section 2.2.1), whereas loudness level is expressed in phon. This explains the difference in absolute values between $\mathrm{PNE}_{\mathrm{loud}}$ and L_N. The average difference between $\mathrm{PNE}_{\mathrm{loud}}$ and L_N was 16 dB. The corresponding difference between L_A and L_N was 18 dB.

It is noteworthy that the regression line for L_A (Fig. 2(a)) falls closely to the sounds with medium low-frequency content (open circles). If only these sounds had been included in the experiment, then L_A would have performed as well as L_N ($R^2 = 0.994$ and 0.995, respectively). This shows the importance of selecting the right set of experimental sounds in order to distinguish between *A*-weighted sound pressure level and loudness level.

3. Experiment 2: annoyance

In Experiment 1, L_{C-A} was found to explain a large part of the variance in perceived loudness not accounted for by L_A. Furthermore, L_N was found to be superior to L_A as an indicator of perceived loudness. The purpose of Experiment 2 was to explore whether these results would hold also for perceived annoyance.

3.1. Method

In Experiment 2, a new group of participants were asked to assess perceived annoyance of the sounds used in Experiment 1. Unlike the loudness instruction used in Experiment 1, the annoyance instruction in Experiment 2 referred to a specific situation: "Imagine that you are sitting in a garden relaxing when you hear the sound." Otherwise, the two experiments were identical; they used the same experimental sounds, the same procedure and the same equipment.

The listeners in Experiment 2 were 31 university students (21 women, 10 men; mean age = 28 years), all with a hearing threshold level [20] below 25 dB in their best ear in the tested frequencies (0.5, 1, 2, 3, 4, 6, and 8 kHz, pure-tone audiometer: Brüel & Kjær, Type 1800). The listeners received course credit for their participation.

3.2. Results and discussion

The consistency of individual data was assessed in the same way as in Experiment 1. All except one participant had test–retest coefficients greater than 0.85 for the pink-noise reference sounds (median = 0.96, range: 0.53–0.99) and all except four participants had coefficients greater than 0.85 for the traffic sounds (median = 0.92, range: 0.62–0.96). Thus, the consistency of individual data was as high as that found in Experiment 1 for loudness. The general trend of the data of the participants with coefficients below 0.85 was not different from the other participants. Therefore, data from all participants were included in the group analyses.

3.2.1. Pink noise equivalent sound level ($\mathrm{PNE}_{\mathrm{annoy}}$)

$\mathrm{PNE}_{\mathrm{annoy}}$-values of annoyance were calculated in the same way as described in connection with Eqs. (1)–(3), but with R_{pn} and R_{traffic} now referring to annoyance rather than to loudness. The fit, R^2, of the psychophysical functions for annoyance, Eq. (1), was excellent, between 0.79 and 0.98 (median = 0.95). The slope, b, of the functions varied between 0.006 and 0.029, median = 0.015.

The inter-individual variability of $\mathrm{PNE}_{\mathrm{annoy}}$ was much greater than for $\mathrm{PNE}_{\mathrm{loud}}$ (Experiment 1). The mean standard deviation across individuals for a given traffic sound was 12.4 dB $\mathrm{PNE}_{\mathrm{annoy}}$ in Experiment 2 (range: 6.7–24.4 dB) compared to 5.4 dB $\mathrm{PNE}_{\mathrm{loud}}$ in Experiment 1. Furthermore, the underlying distributions were negatively skewed in Experiment 2, with the difference between the arithmetic mean and the median $\mathrm{PNE}_{\mathrm{annoy}}$ ranging between -8.7 and -0.16 dB (mean = -2.7 dB). The main reason for the high standard deviation and the skew of the underlying distributions was that some listeners found the pink-noise extremely annoying in comparison to the traffic sounds. This resulted in very low $\text{PNE}_{\text{annoy}}$-values ($<0$ dB in a few cases). Because of the skew distribution of individual $\text{PNE}_{\text{annoy}}$-values, median values were used for summarizing the results across listeners.

3.2.2. *Psychoacoustic relationships for perceived annoyance (PNE_{annoy})*

Fig. 3(a) shows median $\text{PNE}_{\text{annoy}}$-values as a function of L_A. The line shows the linear regression function; the corresponding equation, fit and standard error of the estimate are given in the second row of Table 3.

The general pattern of the data was similar to the pattern found for loudness (cf. Fig. 2(a)). Sounds with high levels of L_{C-A} (filled circles) were more annoying than sounds with medium levels of L_{C-A} (open circles), which in turn were more annoying than sounds with low levels of L_{C-A} (triangles). However, the absolute values of $\text{PNE}_{\text{annoy}}$ were smaller than the corresponding values of PNE_{loud} (compare Figs. 2 and 3). This means that a traffic sound that was found to be approximately equally loud as a given pink-noise level in Experiment 1, was assessed as less annoying than the same pink-noise level in Experiment 2. This is explained by the annoying character of pink noise; several listeners pointed out that the character of the pink noise was more annoying than the character of the road-traffic sounds.

The standard error of the estimate of the linear function in Fig. 3(a) was 3.5 dB $\text{PNE}_{\text{annoy}}$. This error is greater than that found for loudness in Experiment 1 (2.6 dB PNE_{loud}). One reason for this was the poor fit of the three least annoying sounds ($\text{PNE}_{\text{annoy}}<42$ dB). The validity of these $\text{PNE}_{\text{annoy}}$-values may be questioned because they were derived from *extrapolations* of individual psychophysical functions (the lowest pink-noise level was 42 dB). With these three sounds excluded, the standard error of estimate was 3.1 dB.

In order to determine the effect of low-frequency content on perceived annoyance, a standard multiple regression analysis was conducted with $\text{PNE}_{\text{annoy}}$ as the dependent variable and L_A and L_{C-A} as the independent variables. The empirical equation obtained is given in the third row of Table 3, together with the

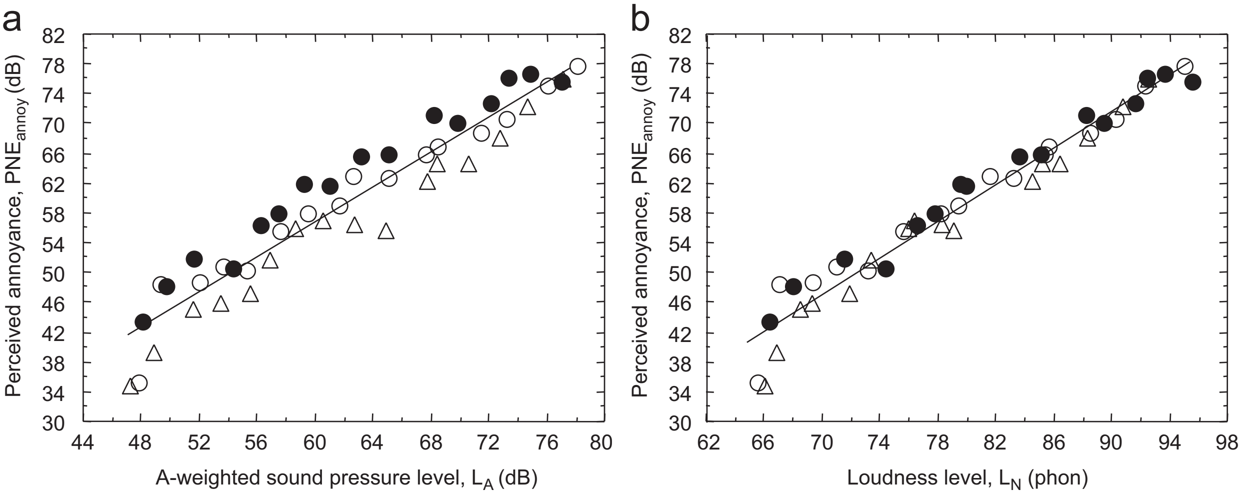

Fig. 3. Perceived annoyance ($\text{PNE}_{\text{annoy}}$) as a function of: (a) A-weighted sound pressure level, L_A, and (b) loudness level, L_N (ISO 532B [11]). ● High, ○ medium and, △ low relative level of low-frequency content.

Table 3
Linear relationships between perceived annoyance ($\text{PNE}_{\text{annoy}}$) and acoustical variables

Empirical relationship	R^2	Standard error (dB $\text{PNE}_{\text{annoy}}$)
$\text{PNE}_{\text{annoy}} = -13.87 + 1.18L_A$	0.91	3.5
$\text{PNE}_{\text{annoy}} = -24.04 + 1.27L_A + 0.46L_{C-A}$	0.96	2.3
$\text{PNE}_{\text{annoy}} = -39.21 + 1.23L_N$	0.95	2.5

model fit (R^2) and the standard error of estimate. All coefficients were significantly greater than zero ($p<0.001$). The coefficients for L_A and L_{C-A} and their ratio ($0.46/1.27 = 0.36$) were similar to the corresponding values for loudness in Experiment 1 ($0.42/1.03 = 0.41$)[1].

Fig. 3(b) shows $\mathrm{PNE}_{\mathrm{annoy}}$ as a function of loudness level, L_N. The line shows the linear regression function; the corresponding equation, fit and standard error of the estimates are given in the fourth row of Table 3. As for loudness in Experiment 1, the standard error of the estimates was only slightly higher for L_N than for the model including both L_A and L_{C-A}, 2.5 and 2.3 dB PNE_{annoy}, respectively (see Table 3).

The regression line for L_A (Fig. 3(a)) falls closely to the sounds with medium low-frequency content (open circles), as also found in Experiment 1. If only these sounds had been included in the experiment, then L_A would have performed as well as L_N, with $R^2 = 0.953$ and 0.954, respectively. Once again, this shows the importance of stimulus selection in the testing of indicators.

4. General discussion

The results showed that sounds with similar A-weighted sound pressure levels were louder and more annoying the greater their relative level of low-frequency content (L_{C-A}). These results agree with the notion that A-weighted sound pressure level overcompensates for the lower sensitivity of the hearing system at low frequencies (e.g., Refs. [1–3]). Conversely, sounds of similar loudness levels [11] were approximately equal in loudness and annoyance irrespective of their low-frequency content. Thus, loudness level was found to be superior to A-weighted sound pressure level as an indicator of short-term loudness and annoyance of road-traffic sounds with a wide variation in low-frequency content.

Several previous experiments on perception of road-traffic sounds have failed to show that A-weighted sound pressure level is inferior to loudness level [7,9]. This was probably because of the lack of systematic selection of experimental sounds in these experiments, which resulted in a set of road-traffic sounds with limited variations in spectral content for sounds of similar sound level. The present results illustrate the consequences of a limited stimulus set: If only traffic sounds with medium relative level of low-frequency content were included in the analyses (open circles in Figs. 2 & 3), then A-weighted sound pressure level would be as successful as loudness level in predicting perceived loudness and annoyance. However, by including several sounds with the same overall sound level that differed in spectral content the indicators were critically tested.

The present results agree with Watts ([10], experiment 3) who found that A-weighting was inferior to loudness level as an indictor of perceived noisiness. Watts filtered road-traffic recordings in order to obtain a set of road-traffic sounds with a wide range of low-frequency content at a constant A-weighted sound pressure level. The present experiments show that Watts's results may be generalized to real, non-manipulated road-traffic sounds.

Because Watts [10] manipulated low-frequency content while keeping other variables constant, his results strongly suggest a causal link between low-frequency content and perception of road-traffic sound. In the present experiments, L_{C-A} was not manipulated independently of other variables. Therefore, one cannot exclude that the observed effect was in fact caused by some "confounding" variable that was also correlated with L_{C-A}.

Inter-correlations between acoustical variables identified one potential confounder, namely the binaural level difference (ΔL), which was moderately correlated with L_{C-A} ($r = -0.47$, Table 1). However, statistical analysis showed that the effect of L_{C-A} on perceived loudness and annoyance remained after controlling for the effect of ΔL (see footnote 1). The other investigated acoustical variables could also be excluded as confounders. Sound-level variability over time was only weakly correlated with L_{C-A}, $r = -0.13$ and 0.04 for L_{A10}–L_{A90} and L_{N10}–L_{N90}, respectively. The two measures of maximum sound level, L_{A10} and L_{N10}, were correlated with L_{C-A}, $r = -0.31$ and -0.15, respectively. However, these variables were almost perfectly correlated with the measures of overall sound level, L_A and L_N. Therefore, they could not have accounted for the variance in perceived loudness or annoyance that was not accounted for by L_A and L_N.

Source identification is another possible confounder. Versfeld and Vos [24] found that heavy vehicles were more annoying than light vehicles at similar A-weighted sound pressure levels, despite similar frequency content. However, several other studies found no effect of source type on annoyance of road-traffic sound

[4,5,8,25]. Carefully listening to the selected excerpts suggested that the variation in low-frequency content was mainly caused by source characteristics and distance to the vehicles. Excerpts with a high relative level of low-frequency content and a *high* overall sound level contained sounds of motorcycles and heavy vehicles at close distance. Excerpts with a high relative level of low-frequency content and a *low* overall sound level contained undifferentiated road-traffic sound from vehicles at further distance. The effect of low-frequency content was found for all sounds, including those at low levels where single sources could not be identified (see Figs. 2 and 3). This speaks against the hypothesis that source identification confounded the observed relationship between low-frequency content and loudness or annoyance.

In summary, the present results suggest that low-frequency content is a determinant of perceived loudness and annoyance of road-traffic sound: A relationship was found and possible confounders were ruled out. Furthermore, the present study quantified the effect of low-frequency content. It was estimated that the change in perceived loudness or annoyance associated with a 1 dB change in L_{C-A} would correspond to approximately a 0.4 dB change in sound level (L_A). Therefore, two road-traffic sounds with the same L_A, one with 0 dB and the other with 15 dB L_{C-A}, would differ in perceived loudness corresponding to a sound level difference of $15 \times 0.4 = 6$ dBA.

Consequently, a corrected sound level for measuring road-traffic sound may be suggested: $L^*_A = L_A + 0.4L_{C-A}$. This correction adds a “penalty” to the sound level depending on its low-frequency content (L_{C-A}). L^*_A may be used as a rule of thumb for *comparing* traffic sounds with different spectral content. Observe, however, that the size of this correction was derived from the present results. One should therefore be cautious about generalizing to other sets of traffic sounds, for instance sounds from larger roads or larger distances to roads than those used in the present experiment.

Previously, Vos [26] and Kjellberg et al. [27] suggested indicators based on both *C*- and *A*-weighted levels for annoyance of impulse sounds and occupational noise. The present results show that such an approach may also be useful for evaluating short-term loudness or annoyance of road-traffic sound. It should, however, be noted that the loudness level (L_N) performed nearly as well as the model based on both *C*- and *A*-weighted sound pressure level in predicting perceived loudness and annoyance (Tables 2 and 3). This speaks in favor of L_N, because it was tested with one free parameter less than the model based on *C*- and *A*-weighted sound pressure level. On the other hand, loudness level is more difficult to measure than *A*- and *C*-weighted sound pressure level. For example, most hand-held sound level meters measure both *A*- and *C*-weighted sound pressure level, but not loudness level. Furthermore, for traffic sounds with typical levels of low-frequency content, *A*-weighted sound pressure level and loudness level were equally good as indicators of perceived loudness and annoyance (open circles in Figs. 2 and 3). Thus, as previously shown for other environmental sounds [23], *A*-weighted sound pressure level is a useful indicator of short-term loudness and annoyance of road-traffic noise without prominent low-frequency content.

5. Conclusions

The main purpose of the present experiments was to determine whether *A*-weighting is a valid indicator of perceived loudness or annoyance of road-traffic sound with wide variation in low-frequency content. The following conclusions were drawn from the results.

(1) *A*-weighted sound pressure level was found to be inferior to loudness level [11] as an indicator of short-term loudness and annoyance of road-traffic sounds with a wide variation in low-frequency content.
(2) Low-frequency content was found to influence the perceived loudness and annoyance of road-traffic sound. In general, sounds with similar *A*-weighted sound pressure levels (L_A) were louder and more annoying the greater their low-frequency content (L_{C-A}).
(3) The change in perceived loudness or annoyance associated with a 1 dB change in L_{C-A} corresponded to approximately a 0.4 dB change in L_A.
(4) Systematic selection of experimental sounds from a “stimulus population” was found to be useful for obtaining non-manipulated environmental sounds with a wide range in acoustical variables critical for the research question. In the present study, the selection method yielded a set of road-traffic sounds with a wide range in both L_A and L_{C-A}.

Acknowledgements

This research was supported by grants to the author from the Institute of Environmental Medicine, Karolinska Institutet and grants to Professor Birgitta Berglund from the Swedish Foundation for Strategic Environmental Research (MISTRA). I would like to thank Ms. Nina Miöen for assistance in testing listeners, Mr. Mikael Andéhn for assistance in data analyses, and Dr. Tomasz Kaczmarek and Dr. Piotr Pekala for computer programming. I would also like to thank Professor Birgitta Berglund, Dr. Bo Schenkman, and two anonymous reviewers for their constructive comments.

References

[1] P.V. Bruel, Is *A*-weighting of noise correct, in: R. Boone (Ed.), *Inter-Noise 2001*, Nederlands Akoestisch Genootschap, Maastricht, The Netherlands, 2001, pp. 1521–1524.

[2] P.D. Schomer, The importance of proper integration of and emphasis on the low-frequency sound energies for environmental noise assessment, *Noise Control Engineering Journal* 52 (2004) 26–39.

[3] A. Kjellberg, M. Goldstein, F. Gamberale, An assessment of dB(A) for predicting loudness and annoyance of noise containing low frequency noise, *Journal of Low Frequency Noise and Vibration* 3 (1984) 10–16.

[4] G.W. Cermak, P.C. Cornillon, Multidimensional analyses of judgments about traffic noise, *Journal of the Acoustical Society of America* 59 (1976) 1412–1420.

[5] G.W. Cermak, Exploratory studies of the relative aversiveness of traffic sounds, *Journal of the Acoustical Society of America* 65 (1979) 112–123.

[6] S. Kuwano, S. Namba, H. Fastl, On the judgment of loudness, noisiness and annoyance with actual and artificial noises, *Journal of Sound and Vibration* 127 (1988) 457–465.

[7] S. Namba, S. Kuwano, Psychological study on Leq as a measure of loudness of various kinds of noises, *Journal of the Acoustical Society of Japan (E)* 5 (1984) 135–148.

[8] S.L. Yaniv, W.F. Danner, J.W. Bauer, Measurement and prediction of annoyance caused by time-varying highway noise, *Journal of the Acoustical Society of America* 72 (1982) 200–207.

[9] G. Watts, P.M. Nelson, The relationship between vehicle noise measures and perceived noisiness, *Journal of Sound and Vibration* 164 (1993) 425–444.

[10] G. Watts, A comparison of noise measures for assessing vehilce noisiness, *Journal of Sound and Vibration* 180 (1995) 493–512.

[11] *ISO, Acoustics—Method for Calculating Loudness Level, ISO 532-1975 (E)*, ISO, Geneva, Switzerland, 1975.

[12] E. Zwicker, H. Fastl, *Psychoacoustics: Facts and Models*, Berlin, Springer, 1990.

[13] J.R. Hassall, Noise measurement techniques, in: C.M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, Acoustical Society of America, Woodbury, NY, 1998, pp. 9.1–9.22.

[14] P. Lundquist, K. Holmberg, U. Landström, Low frequency noise and annoyance in classroom, *Journal of Low Frequnecy Noise, Vibration and Active Control* 19 (2000) 175–182.

[15] J. Vos, *A*- and *C*-weighted sound levels as predictors of the annoyance caused by shooting sounds, for various facade attenuation types, *Journal of the Acoustical Society of America* 113 (2003) 336–347.

[16] S. Kuwano, Temporal aspects in the evaluation of environmental noise, in: D. Cassereau (Ed.), *Inter Noise 2000*, Noise Control Foundation, Poughkeepsie, NY, 2000, pp. 109–119.

[17] E. Pöppel, A hierarchical model of temporal perception, *Trends in Cognitive Sciences* 1 (1997) 56–61.

[18] D. Kahneman, Objective happiness, in: D. Kahneman, E. Diener, N. Schwartz (Eds.), *Well-being: The Foundations of Hedonic Psychology*, Russel Sage Foundation, New York, 1999, pp. 3–25.

[19] B. Berglund, Quality assurance in environmental psychophysics, in: S.J. Bolanowski, G.A. Gescheider (Eds.), *Ratio Scaling of Psychological Magnitudes—In Honor of the Memory of S.S. Stevens*, Erlbaum, Hillsdale, NJ, 1991, pp. 140–162.

[20] ISO, *Acoustics-Standard Reference Zero for the Calibration of Pure-Tone Air Conduction Audiometers*, Geneva, Switzerland, ISO, 1991.

[21] S.S. Stevens, *Psychophysics. Introduction to its Perceptual, Neural and Social Prospects*, New York, Wiley-Interscience, 1975.

[22] L.E. Marks, *Sensory Processes. The New Psychophysics*, New York, Academic Press, 1974.

[23] S. Kuwano, S. Namba, H. Miura, Advantages and disadvantages of *A*-weighted sound pressure level in relation to subjective impression of environmental noises, *Noise Control Engineering Journal* 33 (1989) 107–115.

[24] N.J. Versfeld, J. Vos, Annoyance caused by sounds of wheeled and tracked vehicles, *Journal of Acoustical Society of America* 101 (1997) 2677–2685.

[25] N.J. Versfeld, J. Vos, *A*-weighted sound level as a predictor of the annoyance caused by road traffic consisting of various proportions of light and heavy vehicles, *Journal of Sound and Vibration* 253 (2002) 389–399.

[26] J. Vos, On the annoyance caused by impulse sounds produced by small, medium-large, and large firearms, *Journal of the Acoustical Society of America* 109 (2001) 244–253.

[27] A. Kjellberg, M. Tesarz, K. Holmberg, U. Landström, Evaluation of frequency-weighted sound level measurements for prediction of low-frequency noise annoyance, *Environment International* 23 (1997) 519–527.

[28] B.G. Tabachnick, L.S. Fidell, *Using Multivariate Statistics*, fourth ed., London, Allyn and Bacon, 2001.