Journal of Sound and Vibration (2002) **255**(3), 501–518
doi:10.1006/jsvi.2001.4166, available online at http://www.idealibrary.com on IDEAL®

AFFECTIVE EVALUATIONS OF AND REACTIONS TO EXTERIOR AND INTERIOR VEHICLE AUDITORY QUALITY

D. VÄSTFJÄLL

Department of Applied Acoustics, Chalmers University of Technology, SE-412 96 Göteborg, Sweden. E-mail: daniel@ta.chalmers.se

and Department of Psychology, Göteborg University, Sweden

M.-A. GULBOL AND M. KLEINER

Department of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden

AND

T. GÄRLING

Department of Psychology, Göteborg University, Sweden

(*Received* 8 *February* 2001, *and in final form* 31 *October* 2001)

Affective reactions to and evaluations of auditory stimuli are fundamental components of human perception. In three experiments, participants rated their affective reactions (how pleasant I feel) and preferences for these affective reactions (how much I like the way I feel) as well as affective evaluations (how pleasant the sound is) to interior and exterior binaurally recorded vehicle sounds varying in physical properties. Consistent with previous research, it was found that the orthogonal affect dimensions of valence (unpleasant–pleasant) and arousal or activation (deactivation–activation) discriminated between affective reactions induced by the different qualities of the sounds. Moreover, preference for affective reactions was related to both valence and activation. Affective evaluations (powerful–powerless/passive–active and unpleasant–pleasant) correlated significantly with affective reactions to the same sounds in both within-subjects and between-subjects designs. Standard sound quality metrics derived from the sounds correlated, however, poorly with the affective ratings of interior sounds and only moderately with affective ratings of exterior sounds. Taken together, the results suggest that affect is an important component in product auditory quality optimization.

1. INTRODUCTION

Affective or emotional reactions are fundamental components of human responses to auditory stimuli. Everyday examples of affective reactions to auditory stimuli can easily be found: People may be annoyed by the sound of a car passing by, get tired by the constant fan noise in the office, enjoy the rumbling noise of a motorcycle in a street, be startled by the sudden noise of a door slamming or feel sad by a moving piece of music they hear. Clearly, we react affectively to the auditory environment surrounding us. Moreover, when verbally describing sounds, people consistently use affect-laden words, such as pleasant, tiring, annoying, irritating, happy, and so forth [1].

0022-460X/02/$35.00

Affective reactions to and affective evaluation of auditory stimuli is, however, a largely neglected component of product sound quality development. Sizeable efforts have been made to improve the exterior and interior sound environment of products, and in such attempts specific sound quality metrics have been developed [2]. However, such psychoacoustic metrics do not seem to account for or predict affective responses to auditory stimuli [3, 4]. Moreover, given that individuals react emotionally to sounds, an additional assumption is that people like certain affective states (elation, pleasure, or comfort) and dislike others (annoyance, tiredness, or discomfort). This preference for affective reactions will then guide behavior, in that people approach stimuli that results in preferred affective states, and avoid stimuli that induce disliked affective reactions [5]. Following this line of reasoning, one aim of sound quality research is to optimize product sounds so that they result in preferred affective states induced by the sounds. The current article addresses the fundamental questions of (1) how individuals react to, and affectively evaluate sounds, and (2) how preference is related to affective reactions induced by the sounds.

Traditionally, researchers have sought to establish the underlying perceptual dimension that people rely on when evaluating [6–12] or reacting to sounds [13–15]. However, it *may* be important to distinguish between *evaluations of* and *reaction to* sounds. Typical items in a scale intended to measure perceived sound quality are ratings of sharpness, pleasantness, and annoyance. It can be argued that the rated sharpness of a sound is an evaluation of the sound. A person uses his/her analytic ability to judge the perceptual property sharpness. Annoyance on the other hand, is a mainly affective reaction to a sound. The antecedent to annoyance could of course be cognitive/evaluative ("this sound is very sharp, therefore it is disturbing") and, conversely the state annoyance could be evaluated cognitively ("I am annoyed, probably because this sharp sound is really disturbing"). However, the core of annoyance is an affective reaction. The pleasantness rating is more complex and it could be argued that for such items a confounding exists. Take the example "*this sound is pleasant*" and compare it with "*this sound makes me feel pleasant*". The former refers to a situation where the sound is the object, while the latter refers to a situation where the individual reaction to the sound is the object. Previous research has focused on either affective reactions or evaluations or failed to make the distinction. The current article is concerned with the question of whether affective evaluations (i.e., this sound is pleasant) and affective reactions (this sound makes me feel pleasant) empirically overlap.

Previous research has shown that affective reactions to auditory stimuli may be described by a limited number of underlying dimensions. Björk [16] and Bradley and Lang [17] found that two bipolar dimensions, activation–deactivation and pleasantness–unpleasantness (valence) describe the affective reactions to the quality of natural sounds. The underlying idea is that affective reactions induced by auditory stimuli, or the affective evaluation of the sounds, can be described by a *combination* of the two orthogonal dimensions valence and activation [18, 19]. Figure 1 displays the affect circumplex with the primary dimensions of activation and valence as well as the intermediate dimensions unpleasant activation–pleasant deactivation and pleasant activation–unpleasant deactivation [18]. According to this model, an affective reaction like elation is a combination of pleasantness and high activation (upper right quadrant of Figure 1). An affective reaction like calmness is similar in pleasantness, but low in activation (lower right quadrant). Boredom is a combination of unpleasantness and low activation (lower left quadrant) and distress is a combination of unpleasantness and high activation (upper left quadrant). Thus, the in acoustic research, well-studied affective reaction annoyance would be positioned in the upper left quadrant of Figure 1 [20].

Similar to the two-dimensional view of affective *reactions*, Bisping [2, 3] proposed that affective *evaluations* are fundamental to the perception of interior car sound quality.

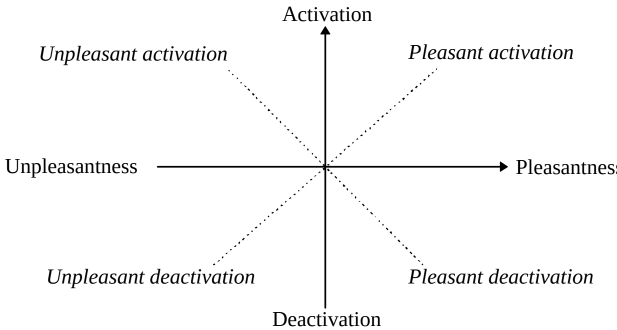

Figure 1. The circumplex model of affect.

Analogous to the dimensions valence and activation, Bisping [3] suggested that *pleasantness* and *powerfulness* of sounds form a two-dimensional perceptual space for the evaluation of car interior sound quality. This two-dimensional space was found to discriminate between different types of interior car sounds. Bisping [3] showed that sounds of luxury cars were positioned in the powerful/pleasant quadrant, while sounds from sporty cars were mainly scattered in the powerful/unpleasant quadrant. The ratings of interior sound from standard middle-size cars were mainly found in the powerless/pleasant quadrant, whereas the powerless/unpleasant quadrant contained sounds from trucks and small cars. Bisping's results thus suggest that the evaluation of sound quality is linked to the two affective dimensions of powerfulness and pleasantness. Further, Bisping's studies show that the sound quality of products may be improved by making the sounds more pleasant and/or powerful. Implicit in such an assumption is that people dislike certain affective states (unpleasant and low/high activation/powerfulness) and like others (pleasant and low/high activation/powerfulness).

Provided that the two dimensions of valence and activation describe the affective reactions to auditory stimuli, this raises the question of how preference is related to these dimensions? In studies of affect-eliciting qualities of environments, Mehrabian and Russell proposed (their "pleasure-arousal hypothesis") that an approach tendency or preference is directly related to valence [5, 21, 22]. The relationship to activation was hypothesized to be inverted U-shaped for a preference maximum that increases with valence.

As may be seen in Figure 2, the pleasure-arousal hypothesis predicts that for affect states similar in unpleasantness, low and moderate activation, unpleasant states will be preferred, for valence-neutral affective states, medium activation will be preferred, and for high valence states, high activation will be preferred over medium and low activation. In short, for unpleasant states, people prefer to feel bored (low activation state) over distressed (high activation state). For pleasant states, people prefer to feel elated (high activation state) over calm (low activation state).

Västfjäll *et al.* [23] found that preference for current mood and emotional reactions were similarly related to valence and activation. Support was also obtained for the particular form of this relationship predicted by Mehrabian and Russell's pleasure-arousal hypothesis [5, 21, 22, 24]. This research thus suggests that participants will hold different preferences for affective reactions caused by auditory stimuli, depending on the degree of activation and valence.

The present study aimed at (1) investigating the relationship between affective reactions and affective evaluations of auditory stimuli, (2) investigating how preferences for affective reactions are related to valence and activation, and (3) assessing the correlation between

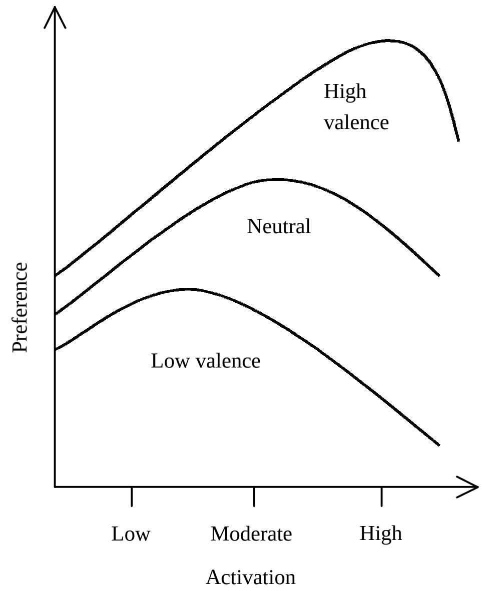

Figure 2. The pleasure-arousal hypothesis (adapted from reference [21]).

affective reaction to sound and sound quality metrics. Testing these hypotheses will contribute to the understanding of how affective perception is related to measurable sound quantities as well as provide insights into the structure of affective reactions.

2. EXPERIMENT

In a within-subjects design, participants listened to 20 interior aircraft sounds and rated their affective reactions and preferences for these reactions as well as their affective evaluations of the sounds.

2.1. METHOD

2.1.1. *Participants*

Twenty undergraduates at Chalmers University of Technology, Göteborg, an equal number of men and women, participated on a voluntary basis. They were compensated with the equivalent of 5 US$. Their mean age was 25·2 yr (S.D. 6·1). All participants had normal hearing as determined by audiogram.

2.1.2. *Measures*

The affect measure consisted of 12 bipolar adjective scales (see reference [25], Experiments 2 and 3 for a more detailed description). Three scales defined by the adjective pairs (translated from the Swedish) dissatisfied–satisified, sad–glad, and depressed–happy were included to tap *valence*. Three other scales defined by the adjective pairs sleepy–awake, dull–peppy, and passive–active were included to tap *activation*. An additional three scales defined by the adjective pairs bored–interested, indifferent–engaged, and pessimistic–optimistic, were included to tap *pleasant activation/unpleasant deactivation*. A final three scales defined by the adjective pairs tense–serene, anxious–calm, and nervous–relaxed were included to tap *unpleasant activation/pleasant deactivation*. The adjective scales defining each dimension were summed (with appropriate signs) and averaged to yield bipolar indices of the four dimensions. The adjective pairs attractive–unattractive, likeable–dislikeable, and "preferred–not preferred to a neutral state" defined three scales included to assess preference. The response format required that participants indicated a number from 10 (defined by left-end adjective) to 90 (defined by right-end adjective) through 50 (neutral). Four different orders of the scales were used with the preference scales always appearing last.

For the affective evaluation ratings, two bipolar adjective scales were used denoted by the adjective passive–activate and unpleasant–pleasant respectively. The same response format was used for these scales.

2.1.3. *Auditory stimuli and presentation*

Sixteen binaural recordings of interior aircraft sounds from different seats in both turboprop and jet aircraft were used. The recordings were made on a TEAC DAT recorder with calibrated Sennheizer KE 4-211-2 microphones at 48 KHz. Based on pilot studies, four additional sounds were created by synthesis–resynthesis of the original sounds by either amplifying the fundamental frequency or the noise spectra. The aim of the sound synthesis was to obtain sounds that would result in a larger variation in the valence and activation dimensions. To accomplish this, four sounds with varying loudness, tonality, sharpness, roughness and fluctuation strength were created. Psychoacoustic metrics were calculated for all sounds using the IDEAS sound quality module on an SGI work station and the HEAD Acoustics Artemis analysis system on a PC. The resulting 20 sounds thus varied considerably in psychoacoustical properties. Range statistics for the 20 sounds were (range for the four synthesized sounds in parenthesis); dBA: 52–81 (50–92), Roughness: 0·4–0·8 (0·2–0·9) asper, Fluctuation strength: 0·19–0·72 (0·08–0·92) vacil; Tonality: 0·10–0·23 (0·01–0·44) Ws; Sharpness 0·77–1·30 (0·45–1·37) acum. The sound stimuli were presented in an acoustically well-damped room over STAX headphones.

2.1.4. *Procedure*

Participants arrived individually to the laboratory for performing ratings of affective reactions to aircraft sounds. Half of the participants first rated their affective reaction when listening to the sound and the other half first indicated their affective evaluations of the same sound. All participants completed both the ratings of affective reactions and evaluation and as a result they listened to the set of sounds twice. For the ratings of affective reaction, participants were asked to check each scale indicating to what degree the adjectives described how they felt when listening to the sounds. For the evaluation ratings, participants were explicitly instructed to rate the quality of the sound, that is to "rate how passive–active (unpleasant–pleasant) the sound is", avoiding rating how they felt when listening to the sound. Participants were also screened for normal hearing after rating the sounds. The procedure took in total approximately 1 h. After finishing, participants were debriefed, compensated and thanked for their participation.

2.2. RESULTS AND DISCUSSION

2.2.1. *Affective evaluations*

The upper graph of Figure 3 shows mean scores of participants' ratings of active–passive and unpleasant–pleasant qualities of the 20 sounds. The intercorrelation

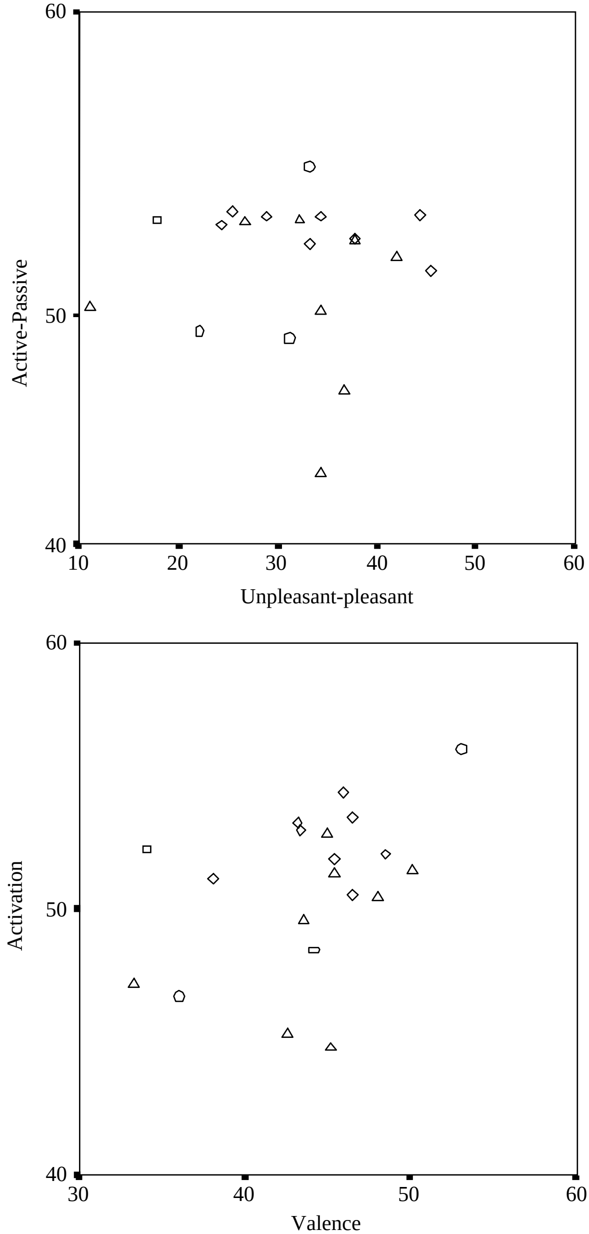

Figure 3. Passive–active ratings plotted versus unpleasant–pleasant ratings for affective evaluations (upper graph) and activation rating plotted versus valence ratings for affective reactions (lower graph) (Experiment 1). □, jet modified; ⬡, turboprop modified; ◇, jet; △, turboproph.

was $-0{\cdot}02$ (ns). As may be seen in Figure 3, the sounds differ in the two evaluative dimensions. This was substantiated by a repeated-measures analysis of variance (ANOVA), $F(2{\cdot}54,\ 47{\cdot}57) = 42{\cdot}31$, $p < 0{\cdot}01$ (after Grenhouse–Geisser correction of the d.o.f.) for active–passive and, $F(2{\cdot}83,\ 51{\cdot}23) = 93{\cdot}14$, $p < 0{\cdot}01$, for unpleasant–pleasant respectively.

2.2.2. *Affective reactions*

The adjective scales tapping valence and activation were summed with appropriate signs and averaged to yield bipolar indices of valence and activation. The six scales measuring pleasant activation–unpleasant deactivation and unpleasant activation–pleasant activation were also summed in the same indices with weights of 0·70 or −0·70 [25]. A preference index was obtained by averaging the ratings on the three adjective scales measuring preference. The intercorrelation between valence and activation was as expected low (0·07 ns).

As may be seen in the lower graph of Figure 3, the affective reactions induced by the aircraft sounds differed on the two dimensions valence and activation. This was substantiated by repeated-measures ANOVAs for activation, $F(2{\cdot}83,\ 49{\cdot}12) = 112{\cdot}33$, $p < 0{\cdot}01$, and valence $F(2{\cdot}66,\ 51{\cdot}73) = 237{\cdot}49$, $p < 0{\cdot}01$.

To check the relation between evaluation and reaction, the evaluation ratings were correlated with the reaction ratings. The correlation for valence and unpleasant–pleasant was 0·72 ($p < 0{\cdot}05$) and for activation and active–passive 0·90 ($p < 0{\cdot}05$). It is thus concluded that both affective evaluations and reactions discriminated between the different sounds, and moreover, there was a substantial overlap between evaluations and reactions.

2.2.3. *Preferences for affective reactions*

To test the hypothesis that both activation and valence are related to preference for affective reactions, the averaged preference index was submitted to regression analysis with the averaged affect indices as independent variables. Neither the quadratic term associated with activation nor the term associated with the interaction between valence and activation reached significance. An additive linear model was fitted for an R^2_{adj} of 0·95, $F(2,\ 19) = 184{\cdot}36$, $p < 0{\cdot}001$, with significant regression weights for both valence ($\beta = 0{\cdot}78$, $t = 19{\cdot}42$, $p < 0{\cdot}001$) and activation ($\beta = -0{\cdot}17$, $t = -3{\cdot}32$, $p < 0{\cdot}05$).

2.2.4. *Sound quality metrics and affective reactions*

To test the relationship between sound quality metrics and the affective reactions, the valence and activation indices were correlated with loudness, fluctuation strength, sharpness, roughness and tonality. For the activation index, none of these indices exhibited a significant correlation. The valence index was related to loudness (0·42, $p < 0{\cdot}05$). These results are in agreement with earlier findings [4].

In sum, Experiment 1 showed that affective reactions and evaluations are overlapping constructs. Further, preference for affective reactions were related to the fundamental affect dimensions valence and activation.

3. EXPERIMENT

A possible explanation for the overlap between affective reactions and evaluations in Experiment 1 is that a within-subjects design was used. Experiment 2 was therefore conducted to replicate Experiment 1 with a between-subjects design.

3.1. METHOD

3.1.1. *Participants*

Forty eight undergraduates at Chalmers University of Technology, Göteborg, 28 men and 20 women, participated on a voluntary basis. They were compensated with the equivalent of US$5. Their mean age was 22·8 yr (S.D. 4·7).

3.1.2. *Measures*

The affective reactions measure consisted of the 12 bipolar adjective scales used in Experiment 1. An additional three items measured preference for affective reactions, also taken from Experiment 1. The affect scales defining each dimension were summed (with appropriate signs) and averaged to yield bipolar indices of the four dimensions.

As in Experiment 1, one group of participants rated affective evaluations on two bipolar adjective scales denoted by the adjective passive–activate and unpleasant–pleasant. In addition, another group of participants made ratings rated on a Swedish translation of Bispings [2, 3] evaluation scales of powerful–powerless and pleasant–unpleasant.

3.1.3. *Auditory stimuli and presentation*

Sixteen of the binaural recordings of interior aircraft sounds from Experiment 1 were used. The sounds were presented in an acoustically well-damped room over STAX headphones.

3.1.4. *Procedure*

Participants served individually. They were randomly assigned to either rate their affective reaction and preferences ($n = 16$), indicate their affective evaluation on the active–passive and unpleasant–pleasant scales ($n = 16$) or on the powerful–powerless and unpleasant–pleasant scales ($n = 16$). Upon arrival to the laboratory, participants were first instructed on how to use the scales and equipment. Participants then listened and rated two trial sounds. After that they rated the 16 sounds. Four random orders were used for each condition. After participants had rated the sounds, they were debriefed, compensated, and thanked for their participation.

3.2. RESULTS AND DISCUSSION

3.2.1. *Affective evaluations*

The upper graph in Figure 4 shows participants' mean ratings of passive–active and unpleasant–pleasant qualities of the 16 sounds. The correlation between the two dimensions was 0·09. As may be seen, the sounds differed in these two dimensions. A repeated–measures ANOVA substantiated this, $F(2{\cdot}45, 49{\cdot}55) = 29{\cdot}14$, $p < 0{\cdot}01$ for active–passive and $F(2{\cdot}64, 52{\cdot}32) = 32{\cdot}77$, $p < 0{\cdot}01$, (after Grenhouse–Geisser correction of the d.o.f.) for unpleasant–pleasant respectively.

The middle graph of Figure 4 displays the mean ratings by the second group of participants of the powerful–powerless and unpleasant–pleasant qualities. The intercorrelation for the two dimensions was $-0{\cdot}13$. Also, for these two dimensions, the mean ratings of the different sounds differed significantly, $F(2{\cdot}22, 55{\cdot}86) = 55{\cdot}92$, $p < 0{\cdot}01$, for powerful–powerless, and $F(3{\cdot}11, 47{\cdot}71) = 72{\cdot}28$, $p < 0{\cdot}01$, for unpleasant–pleasant respectively.

The active–passive ratings by the first group correlated 0·81 ($p > 0{\cdot}05$) with the powerful–powerless ratings of the other group. The unpleasant–pleasant ratings by the first group correlated 0·92 ($p > 0{\cdot}05$) with the ratings by the second group.

3.2.2 *Affective reactions*

As may be seen in the lower graph of Figure 4, the affective reactions induced by the aircraft sounds differed on activation and valence, $F(2{\cdot}16, 47{\cdot}50) = 78{\cdot}30$, $p < 0{\cdot}01$, and

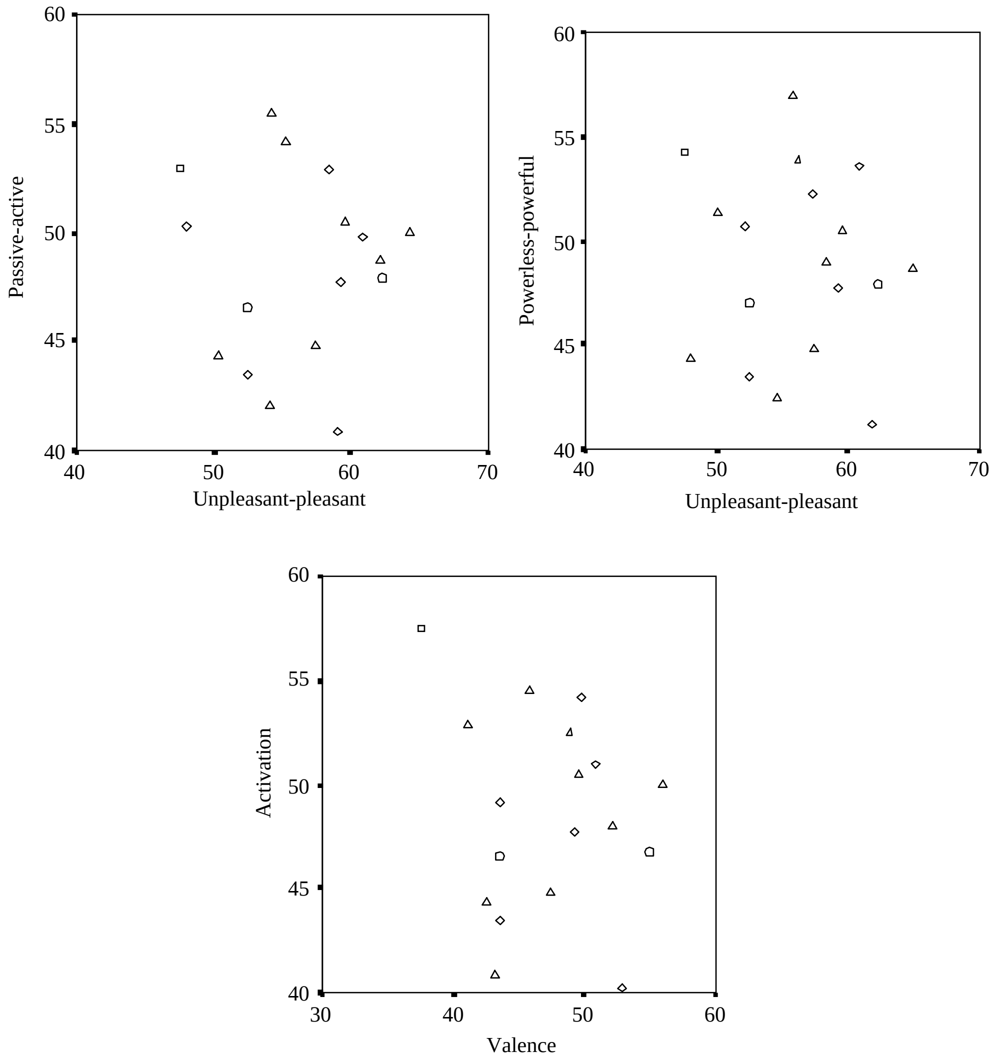

Figure 4. Passive–active ratings plotted versus unpleasant–pleasant ratings (upper graph), powerless–powerful ratings plotted versus unpleasant–pleasant ratings (middle graph) for affective evaluations, and activation ratings plotted versus valence ratings for affective reactions (lower graph) (Experiment 2). □, jet modified; ⬡, turboprop modified; ◇, jet; △, turboproph.

$F(2{\cdot}91,\ 43{\cdot}12) = 29{\cdot}12$, $p < 0{\cdot}01$ respectively. The intercorrelation between valence and activation was 0·11. Ratings of activation correlated 0·90 ($p < 0{\cdot}05$) with active–passive evaluations and 0·76 ($p < 0{\cdot}05$) with powerful–powerless evaluations. Valence correlated 0·91 ($p < 0{\cdot}05$) with unpleasant–pleasant evaluations made by the first group and 0·89 ($p < 0{\cdot}05$) with unpleasant–pleasant evaluations by the second group.

3.2.3. *Preferences for affective reactions*

Participants' ratings of preferences for affective reactions were submitted to multiple regression analyses. As in Experiment 1, neither the quadratic term of activation nor the interaction between valence and activation reached significance. However, a linear additive regression equation was fitted for $R^2_{adj} = 0{\cdot}72$, $F(2, 17) = 22{\cdot}28$, $p < 0{\cdot}01$, indicating that both valence ($\beta = 0{\cdot}64$, $t = 4{\cdot}64$, $p < 0{\cdot}001$) and activation ($\beta = -0{\cdot}39$, $t = -2{\cdot}92$, $p < 0{\cdot}02$) contributed reliably.

In summary, Experiment 2 showed that the overlap between affective evaluations and reactions remained when a between-subjects design was used. Further, affective evaluations of active–passive and powerful–powerless qualities of the sounds overlapped in addition to the expected overlap between unpleasant–pleasant ratings by the two evaluation groups. Experiment 2 showed that both valence and activation determined preferences for affective reactions, but failed to find a quadratic term of activation and an interaction of valence and activation as expected by the pleasure–arousal hypothesis [21]. A possible explanation for this is that the set of sounds used in Experiments 1 and 2 showed a restricted variation in the activation dimension. Previous research has indicated that such a restriction of range may result in failure to fit the quadratic and interaction term [21, 23].

4. EXPERIMENT

Experiments 1 and 2 used relatively stationary and static interior aircraft sounds. To increase variation in the ratings, Experiment 3 employed recordings of 12 time-varying exterior vehicle sounds. In a within-subjects design, participants first rated their affective evaluations and then their affective reactions as well as preference to the twelve sounds.

4.1. METHOD

4.1.1. *Participants*

Forty two undergraduates at Chalmers Univeristy of Technology, Göteborg, 12 female and 30 male, participated on a voluntary basis. Their mean age was 26·2 yr (S.D. = 4·5). They were compensated with the equivalence of USD$6 for their participation.

4.1.2. *Measures*

The affect measure consisted of two bipolar scales each defined by adjective pairs found in previous research (see references [25, 26] for a detailed overview of the development of these scales) to tap valence and activation respectively. The adjective pairs defining the activation scale were sleepy–awake, dull–peppy, and passive–active, and those defining the valence scale displeased–pleased, sad–glad, and depressed–happy. An additional scale was included to measure preference for the affective reaction. This scale was defined by the adjectives attractive, likeable, and "preferred relative to a neutral state" [23]. Beneath each set of adjectives typed on a single page, two endpoints and a middle point defined by the numbers 10, 50, and 90 were typed in boxes from left to right. In between there were two open boxes. Participants were asked to let the three adjective pairs define each scale and to write an appropriate number in the open boxes (11–49 or 51–89) or to cross one of the boxes with numbers. They made ratings on the preference scale by indicating a number between 10 (not at all) to 90 (very much).

Given the substantial overlap between the passive–active and powerless–powerful evaluation scales, either scale could be used. However, the powerless–powerful and unpleasant–pleasant scales were originally developed for the assessment of vehicle sound quality and may thus be more suitable for the assessment of affective qualities of vehicle sounds [3]. For this reason, the powerful–powerless and unpleasant–pleasant scale were chosen. A standard semantic differential 7-point response format was used where powerless (unpleasant) defined the left endpoint and powerful (pleasant) defined the right endpoint [5, 12].

4.1.3. *Auditory stimuli and presentation*

Twelve binaurally recorded exterior car sounds were used. The set of sounds was obtained from six small and medium sized (1300–2000 cm^3) four-cylinder diesel and gasoline cars. The sounds were binaurally recorded under two conditions, either wide open throttle acceleration from 1000 to 3000 r.p.m. with a constant acceleration of 100 r.p.m./s, and throttle pulses. The sounds were replayed to the participants through Sennheiser HD 414 headphones. Psychoacoustic metrics were derived for all recordings. The range statistics for the stimuli dBA: 60–101, Roughness: 0·10–0·94 asper, Fluctuation strength: 0·06–0·90 vacil; Tonality: 0·07–0·69 Ws; and Sharpness 0·61–1·12 acum. Thus, in comparison to Experiments 1 and 2, the sound stimuli in Experiment 3 varied considerably more in measurable psychoacoutic quantities. In addition, the stimuli in Experiment 3 were time-varying sounds, whereas in Experiments 1 and 2 stationary stimuli were used.

4.1.4. *Procedure*

Groups of four to eight participants were tested each time. A male experimenter first instructed the participants on how to use the scales and equipment. Participants then listened to four practice sounds, chosen to cover the variation in the stimuli. After being familiarized with the sounds and scales, participants rated the 12 sounds. For each sound, participants first rated their evaluations on the powerful–powerless and unpleasant–pleasant scales. Additional unrelated questions concerning the sounds were also asked. After this participants rated their affective reactions and preferences for these reactions. When finished, the next sound was replayed. Different orders were used for the different groups of participants. After finishing rating all sounds, participants were debriefed, compensated and thanked for their participation.

4.2. RESULTS AND DISCUSSION

4.2.1. *Affective evaluations*

The upper graph in Figure 5 displays mean ratings of powerful–powerless and unpleasant–pleasant qualities. The intercorrelation for the two dimensions was $-0{\cdot}13$. In repeated–measures ANOVAs the mean ratings of the different sounds differed significantly for both dimensions, $F(2{\cdot}71, 127{\cdot}28) = 24{\cdot}77$, $p < 0{\cdot}01$, for powerful–powerless, and $F(2{\cdot}44, 129{\cdot}32) = 13{\cdot}89$, $p < 0{\cdot}01$, for unpleasant–pleasant respectively (after Grenhousse–Geisser correction of the d.o.f.).

4.2.2. *Affective reactions*

As may be seen in the lower graph of Figure 5, the affective reactions differed on the two dimensions valence and activation. This was substantiated by additional repeated-measures ANOVAs for activation $F(2{\cdot}91, 169{\cdot}42) = 23{\cdot}92$, $p < 0{\cdot}01$, and valence $F(2{\cdot}89, 178{\cdot}91) = 5{\cdot}82$, $p < 0{\cdot}05$. The intercorrelation between valence and activation was 0·10.

The correlation between valence and unpleasant–pleasant was 0·91 ($p < 0{\cdot}05$) and the correlation between activation and powerless–powerful was 0·78 ($p < 0{\cdot}05$).

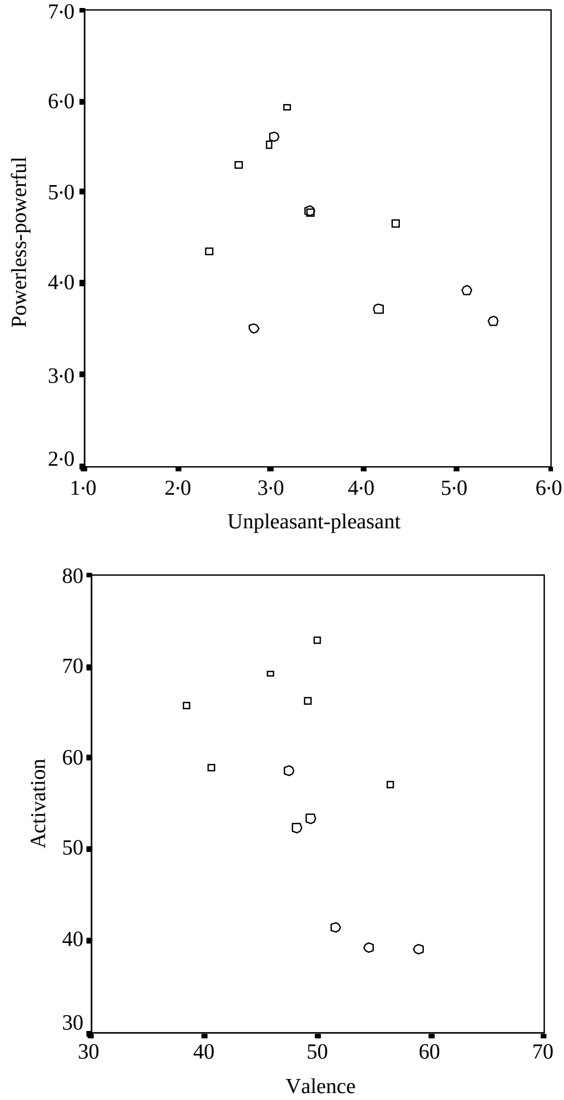

Figure 5. Passive–active ratings plotted versus unpleasant–pleasant ratings for affective evaluations (upper graph) and activation ratings plotted versus valence ratings for affective reactions (lower graph) (Experiment 3). □, Trottle pulse; ⬡, accelerating 1000-3000 rpm.

4.2.3. *Preferences for affective reactions*

Fitting a linear additive model yielded a R^2_{adj} of 0·90, $F(2, 10) = 50{\cdot}29$, $p < 0{\cdot}01$. Only valence ($\beta = 0{\cdot}92, t = 9{\cdot}32, p < 0{\cdot}001$) contributed reliably whereas activation did not reach significance ($\beta = -0{\cdot}12$, $t = -1{\cdot}36$, $p > 0{\cdot}05$). However, when adding the quadratic term for activation and the valence activation interaction, the R^2_{adj} increased to 0·95, $F(2, 10) = 68{\cdot}47$, $p < 0{\cdot}01$, where *valence* ($\beta = 0{\cdot}63$, $t = 6{\cdot}14$, $p < 0{\cdot}001$), a^2 ($\beta = -0{\cdot}22$, $t = 2{\cdot}96$, $p < 0{\cdot}001$) and *axv* ($\beta = 0{\cdot}34$, $t = 3{\cdot}88$, $p < 0{\cdot}001$) all contributed significantly. Figure 6 displays the plot of model-derived preference against observed preference. As may be seen, no systematic deviations are discernable.

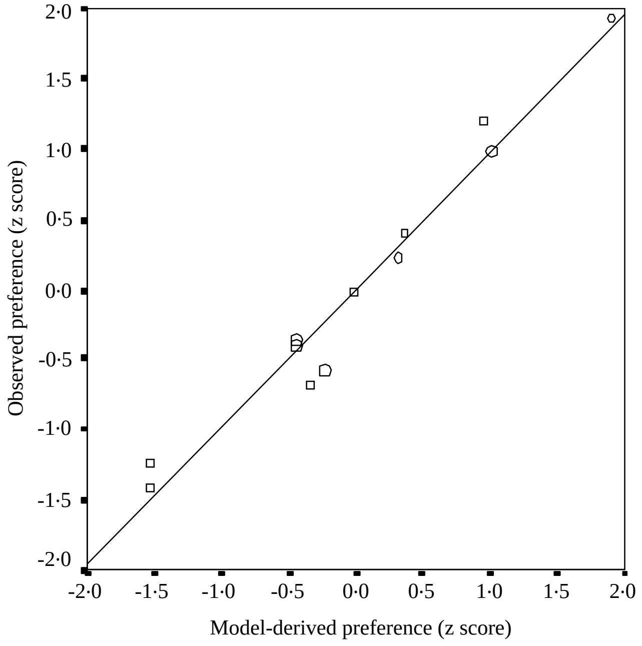

Figure 6. Observed preference plotted versus model–derived preference (Experiment 3). □, Trottle pulse; ⬡, accelerating ——, total population.

4.2.4. *Sound quality metrics and affective reactions*

The sound quality metrics (loudness, sharpness, roughness, fluctuation strength, and tonality) derived from the sounds were entered into multiple regression analyses with either the mean valence or mean activation index as a dependent variable. To improve the predictive power, two separate regression analyses were performed on the valence and activation indices rather than the preference variable. For the activation index, tonality, sharpness, and roughness correlated significantly at $p < 0{\cdot}05$. A subsequent regression analysis gave an R^2_{adj} of 0·66, $F(1,\ 11) = 10{\cdot}75$, $p < 0{\cdot}001$, where the beta weights for roughness ($\beta = 0{\cdot}56$, $t = 3{\cdot}02$, $p < 0{\cdot}02$) and tonality ($\beta = -0{\cdot}55$, $t = 2{\cdot}98$, $p < 0{\cdot}02$) were significant. For the valence index only loudness contributed significantly ($\beta = -0{\cdot}66$, $t = 3{\cdot}02$, $p < 0{\cdot}02$) for a R^2_{adj} of 0·57, $F(1,\ 11) = 14{\cdot}12$, $p < 0{\cdot}001$.

Even if the majority of variance in preference ratings was accounted for by the valence and activation indices, a model directly assessing the relationship between preference and sound quality metrics was tested. A regression analysis with the preference ratings as the dependent variable, and loudness, roughness, and tonality as independent variables, was performed. The analysis showed that only loudness ($\beta = -0{\cdot}52$, $t = 2{\cdot}69$, $p < 0{\cdot}05$) and tonality ($\beta = -0{\cdot}31$, $t = 2{\cdot}13$, $p < 0{\cdot}05$) contributed significantly for an R^2_{adj} of 0·61, $F(1,\ 11) = 9{\cdot}26$, $p < 0{\cdot}01$.

In sum, Experiment 3 showed for a different set of sounds that affective reactions and evaluations share considerable variance. In agreement with previous research, preferences for affective reactions induced by the sounds were related to activation and valence [21]. With increased variation in activation ratings, support was also obtained for the specific form of the pleasure–arousal hypothesis with preference proportional to valence and related to activation through an inverted U-shaped function with a maximum increasing with

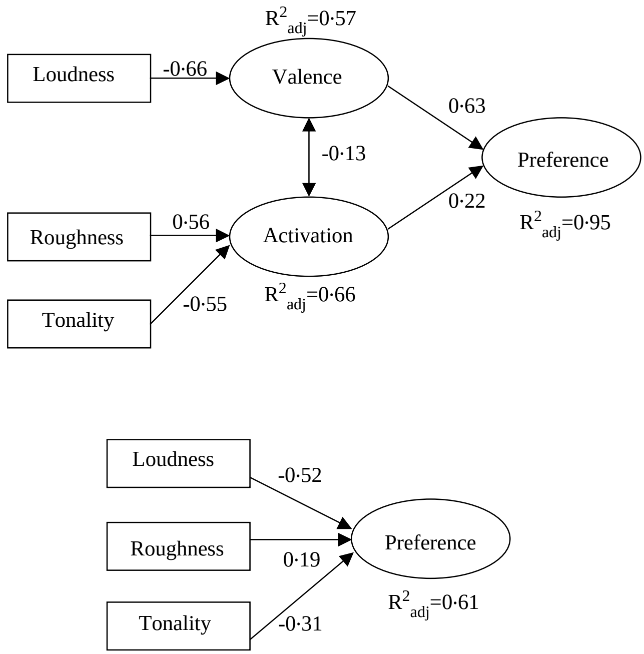

Figure 7. Regression coefficients for the combined regression analyses of preference, affective reactions and psychoacoustic correlates in Experiment 3 (standardized beta weights and explained variance are shown). The upper graph shows the full model, where preference is mediated by valence and activation. The lower graph shows the reduced model, where preference is directly related to psychoacoustic correlates.

valence [23]. Thus preference was, as predicted, mediated by the valence and activation indices. Further, valence and activation significantly correlated with physical characteristics of the sound stimuli. It was also shown that preference was directly related to loudness and tonality, with liking increasing with decreasing loudness and tonality. These relationships are summarized in Figure 7.

5. GENERAL DISCUSSION

Taken together, the three studies provide support for the fact that affective reactions to auditory stimuli could be discriminated by the two fundamental affect dimensions of valence and activation. Different sounds give rise to single affect states characterized by unpleasant and pleasant, deactivated or activated feelings [19]. Provided that affective reactions can differentiate between different auditory stimuli, affect has an important place in product sound quality development. If the aim of product sound quality is to optimize the sound for the user, by minimizing negative responses and maximizing positive response to the sound, affect has to be acknowledged as an integral part of user evaluation and reaction. However, simply mapping affective reactions to auditory stimuli is not enough. One also needs to know what affective state is desirable, that what are the users' preferences are for affective reactions? An account of preferences for affective reactions, such as the pleasure–arousal hypothesis, suggests that people prefer pleasant states to unpleasant states [21]. However, for an affective state similar in pleasantness, people will differ in their preferences depending on the degree of activation (i.e., people will prefer feeling elated over feeling content or calm). This is also the case for affective states similar in unpleasantness (i.e., people will prefer to feel bored over sad and distressed). The pleasure–arousal hypothesis thus suggests that both valence and activation are important for preference for affective reactions. This was substantiated in Experiments 1 and 2, although the results suggested that preference was linearly related to both activation and valence. Support was, however, not found for the specific form of the pleasure-arousal hypothesis as depicted in Figure 2. A possible explanation for this is that the affective reactions induced by the interior aircraft sounds did not vary much in the activation dimension. For this reason and to extend the validity of the results, a different set of sounds were used in Experiment 3, namely binaurally recorded, time-varying exterior vehicle sounds. The use of this new type of sounds induced a larger variation in the activation ratings. Support was then obtained for the fact that the preference for affective reactions to auditory stimuli was related to valence and activation as specified by the pleasure-arousal hypothesis. Still it should be noted that activation, in addition to valence, contributed reliably in determining preferences in all three experiments.

The present research also tested the hypothesis that affective *evaluations* may be different from affective reactions [2]. The reasoning behind this is the idea that affective reactions concern peoples feelings of, for instance, pleasantness with regard to the sound. Affective evaluations, on the other hand, instead concern sensory evaluation of the fact that a sound may be pleasant, without inducing pleasant feelings in the listener. If there is a difference between these potentially different concepts, both research and practical applications using subjective rating scales would need to acknowledge that. For instance, the behavioral consequence of noise exposure may be very different if people actually feel very annoyed by the sound, as compared to only perceive the sound as annoying. Theoretically, the judgmental process for the two types of ratings differ: affective reactions to auditory stimuli entail a reference to the self as reacting to the sound, whereas affective evaluations concern an analysis of the affective qualities of the sound. In the three experiments these constructs overlapped, where affective reactions (*I* feel unpleasant) were significantly positively related to the evaluation (*the sound* is unpleasant). Given that this empirical and conceptual overlap exists, what should be assessed when studying affect and sound quality? It seems that affective evaluations are linked to a cognitive, analytic process where people consciously assess the quality of a sound, whereas an affective reaction may, but need not, be a consequence of the sound environment without the person knowing it. Of course, an affective reaction can also be more or less cognitively salient. An affective reaction is thus related to what Russell and Feldman–Barret [19] termed *core affect*. Core affects are cognitively accessible elements that are present in any type of affective reaction. Further, core affects need not be directed at a specific object. Affective reactions and evaluations may thus both be important for sound quality assessment. However, in many cases the quality of the sound is assessed by having participants or potential users rate perceptual attributes such as loud, rough, soft, clear, and so forth. In those cases, the inclusion of ratings of affective evaluations is an alternative. Other research and development of product sound quality have investigated reactions to product sounds by having participants rate how annoyed or disturbed they are by the sounds. In such cases ratings of affective reactions of valence and activation would be desirable since these dimensions cover a significant portion of affective states, and thus more information beyond annoyance would be gained.

One important goal of sound quality research is to derive objective or psychoacoustic measures that account for people's reactions to the sound. Today several sound quality metrics such as instantaneous and overall loudness, sharpness, roughness, tonality, and fluctuation strength exist [27]. However, they correlate poorly with affective reactions [4]. In Experiment 1, it was found that none of the above-mentioned sound quality metrics were significantly related to activation ratings, and only loudness was related to valence ratings. In Experiment 3, tonality and roughness were, however, significantly related to activation, and overall loudness to valence ratings. This may be due to the fact that activation and valence ratings varied to a larger extent. The auditory stimuli in Experiment 3 were strongly time-varying sounds, either gradually increasing or pulses. The derived sound quality metrics are not well suited for this type of time-varying signals since they only give a rough estimate summed over time. Nevertheless, affective reactions were to some extent related to these metrics. This may be due to the fact that the affective ratings are an average of the affect experienced over the whole time course of the auditory stimuli. Participants were asked to report their affective reactions after listening to each sound. In some way, participants then must average their affective reaction and report a single value for how activated and how unpleasant–pleasant they feel [28]. Consequently, for time-varying sounds it may thus be advantageous to have participants continuously rate their experienced affect on either one dimension of valence or activation [28–30] or in a two-dimensional space [31].

In Experiment 3 it was also found that participants preference ratings were directly related to sound quality metrics, with preference or liking decreasing for increased loudness and tonality. This finding replicates previous research that used similar types of preference ratings [27]. However, compared to a model with preference mediated by valence and activation considerably less variance in the preference ratings could be explained (95% versus 61%). Also, the reduced model only showed a modest, non-significant relationship between preference and roughness. In contrast, the model including valence and activation showed that roughness was significantly positively related to activation, and that activation in turn was significantly related to preference. This points to the benefits of simultaneously assessing subjective reactions in terms of valence, activation, and preference, rather than only preference.

Taken together, affective reactions were not fully accounted for by existing sound quality metrics, and future research should therefore further investigate physical determinants of valence and activation. Moreover, an important step, not only for affective reactions to sound qualities, is to develop sound quality metrics that take into account time variation as may be found for the exterior sound quality of vehicles.

Summing up, in agreement with previous research the current research has shown that affective reactions to and preferences for these affective reactions as well as evaluations of the quality of auditory stimulus are related to unpleasantness–pleasantness and activation–deactivation/powerful–powerless dimensions [3, 5, 16, 17, 18, 32]. Future research should investigate affective reactions to other auditory stimuli, including sound environments with multiple sound sources. Moreover, other assessment methods such as continuous ratings and psychophysiological measures should be used to validate existing methods.

ACKNOWLEDGMENTS

This research was financially supported by NFFP grant no. 232 given to Mendel Kleiner from the Swedish National Aviation Research Program and by the EC program SVEN. We thank Vincent Rioux and Börje Wijk for their assistance in constructing the stimuli and setting up the equipment used to present them.

REFERENCES

1. S. NAMBA 1994 *Inter-noise* **94**, 171–218. Noise—quantity and quality.
2. R. BISPING 1995 *SAE Paper*, *Vol*. 951284, 1203–1209. Emotional effect of car interior sounds: pleasantness and power and their relation to acoustic key features.

3. R. BISPING 1997 *Acta Acustica* **83**, 813–818. Car interior sound quality: experimental analysis by synthesis.
4. D. VÄSTFJÄLL, M. KLEINER and T. GÄRLING 2001 *Report* 02-01, *Department of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.* Affective reactions and cognitive evaluations of interior aircraft sound quality.
5. A. MEHRABIAN and J. A. RUSSELL 1974 *An Approach to Environmental Psychology*. Cambridge, MA: MIT Press.
6. J. A. BALLAS 1993 *Journal of Experimental Psychology: Human Perception and Performance* **19**, 250–267. Common factors in the identification of an assortment of brief everyday sounds.
7. A. GABRIELSSON and H. SJÖGREN 1979 *Journal of the Acoustical Society of America* **65**, 1019–1033. Perceived sound quality of sound reproducing systems.
8. J. M. GREY 1977 *Journal of the Acoustical Society of America* **61**, 1270–1277. Multidimensional perceptual scaling of musical timbres.
9. R. GUSKI 1997 *Acta Acustica* **83**, 765–774. Psychological methods for evaluating sound quality and assessing acoustic information.
10. G. VON BISMARK 1974 *Acustica* **30**, 159–172. Sharpness as an attribute of the timbre of steady sounds.
11. S. MCADAMS, S. WINSBERG, S. DONNADIEU, G. DE SOETE and J. KRIMPHOFF 1995 *Psychological Research* **58**, 177–192. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes.
12. C. E. OSGOOD, G. J. SUCI and P. TANNENBAUM 1957 *The Measurement of Meaning*. Urbana: University of Illinois Press.
13. S. NAMBA, S. KUWANO, T. HASHIMOTO, B. BERGLUND, Z. DA RUI, A. SCHICK, H. HOEGE and M. FLORENTINE 1991 *Journal of the Acoustical Society of Japan* (*E*) **12**, 19–29. Verbal expression of emotional impression of sound: a cross-cultural study.
14. L. N. SOLOMON 1958 *Journal of the Acoustical Society of America* **30**, 421–425. A semantic approach to the perception of complex sounds.
15. L. N. SOLOMON 1959 *Journal of the Acoustical Society of America* **31**, 986–990. Semantic reactions to systematically varied sounds.
16. E. A. BJÖRK 1985 *Acustica* **57**, 185–188. The perceived quality of natural sounds.
17. M. M. BRADLEY and P. J. LANG 2000 *Psychophysiology* **37**, 204–215. Affective reactions to acoustic stimuli.
18. J. A RUSSELL 1980 *Journal of Personality and Social Psychology* **39**, 1161–1178. The circumplex model of affect.
19. J. A RUSSELL and L. FELDMAN–BARRETT 1999 *Journal of Personality and Social Psychology* **76**, 805–819. Core affect, prototypical episodes, and other things called emotion: Dissecting the elephant.
20. N. A. REMINGTON, R. F. FABRIGAR and P. S VISSER 2000 *Journal of Personality and Social Psychology* **79**, 266–300. Reexamining the circumplex model of affect.
21. J. A RUSSELL and A. MEHRABIAN 1978 *Environment and Behavior* **10**, 355–387. Approach-avoidance and affiliation as functions of the emotion-eliciting quality of an environment.
22. A. MEHRABIAN, C. WIHARDJA and E. LJUNGGREN 1996 *Genetic, Social and General Psychology Monographs* **123**, 461–478. Emotional correlates of preferences for situation-activity combinations in everyday life.
23. D. VÄSTFJÄLL, T. GÄRLING and M. KLEINER 2001 *Journal of Happiness Studies* **2**, 337–354. How do I like feeling this way? A core affect account of preference for current mood.
24. A. N. KLUGER, S. LEWINSOHN and J. R. AIELLO 1994 *Organizational Behavior and Human Decision Processes* **60**, 276–299. The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal.
25. D. VÄSTFJÄLL, M. FRIMAN, T. GÄRLING and M. KLEINER 2001 *Scandinavian Journal of Psychology* **43**, 19–31. The measurement of core affect: a Swedish self-report measure derived from the affect circumplex.
26. D. VÄSTFJÄLL and T. GÄRLING 2001 *Göteborg Psychological Reports*, Vol. 9 (1), *Department of Psychology, Göteborg University, Sweden.* Development and validation of short measures of valence and activation: the Swedish core affect scales (SCAS) II.
27. E. ZWICKER and H. FASTL 1999 *Psychoacoustics—Facts and models*. Heidelberg: Springer-Verlag, second edition.

28. C. A. Schreiber and D. Kahneman 2000 *Journal of Experimental Psychology*: *General* **129**, 27–42. Determinants of the remembered utility of aversive sounds.
29. S. Kuwano and S. Namba 1985 *Psychological Research* **47**, 27–37. Continuous judgment of level-fluctuating sounds and the relationship with overall loudness and instantaneous loudness.
30. P. Susini and S. Mcadams 2000 *Acta Acustica* **85**, 515–525. Psychophysical validation of a proprioceptive device by cross-modal matching of loudness.
31. E. Schubert 1999 *Australian Journal of Psychology* **51**, 154–165. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space.
32. M. M. Bradley and P. J. Lang 1999 *Technical Report*, Vol. B-2, *The Center for Research in Psychophysiology*, *University of Florida*. International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings.