EURONOISE 2006
30 May - 1 June 2006 Tampere, Finland

VISUAL BIAS IN SUBJECTIVE ASSESSMENTS OF AUTOMOTIVE SOUNDS

Wolfgang Ellermeier[1] and Søren Vase Legarth[1,2]

[1]Sound Quality Research Unit, Dep. of Acoustics, Aalborg University
Fredrik Bajers Vej 7 B-5, DK-9220 Aalborg Ø., Denmark
we@acoustics.dk

[2]DELTA Danish Electronics, Light and Acoustics, Hørsholm, Denmark

ABSTRACT

In order to evaluate how strong the influence of visual input on sound quality evaluation may be, a naive sample of 20 participants was asked to judge interior automotive sound recordings while simultaneously being exposed to pictures of cars. Twenty-two recordings of second-gear acceleration presented via headphones were combined with an equal number of full-screen color displays of car models selected to exert either a positive, or a negative bias on ratings of the 'powerfulness' of the sounds. After having been presented with the ensuing 44 combinations, subjects also rated the sounds alone, as well as the impression of powerfulness conveyed by the pictures presented in isolation. It turned out that concurrent, task-irrelevant presentation of biasing pictures exerted a strong, and statistically significant influence on the participant's ratings of powerfulness. The variance in responses accounted for by the pictures was roughly one fifth of the variance due to the sounds themselves. In physical terms, the picture manipulation shifted ratings of powerfulness by what a change in overall sound-pressure level of approximately 2-3 decibels would accomplish.

1 INTRODUCTION

Audio-visual interaction encounters growing interest in basic research (for a review, see [1]), but relatively little is known about how visual input influences perceived sound quality. Laboratory research on 'soundscapes' [2], or on the effects of environmental noise [3] typically employed rich visual scenes to accompany the auditory input, but the effects found have often been complex, sometimes providing evidence for positive, other times for negative effects on the judgments of the sounds of interest. More definite answers come from research on the perceived quality of combined television/hifi systems: Here the influence of (non-focal) video quality on perceived audio quality was found to be much higher than the (reverse) influence of audio quality on ratings of video quality [4]. Thus while there is some evidence for audio-visual interaction in judgments of preference, or overall (audio) quality, it remains to be shown that specific sound quality attributes, that play a role in product sound evaluation, for example, are susceptible to interference from visual input.

Therefore, a 'demonstrator' experiment was set up to show such an effect while maximizing the chance for its occurrence: Automotive sounds were chosen as the product sound material to study, and a relatively gullible, 'connotative' (i.e. non-sensory) attribute, the *powerfulness* of the (accelerating car) sound was selected. By presenting pictures to generate either positive or negative expectations regarding the 'power' of the sound to be heard, an attempt was made to effectively 'bias' the sound quality judgments of naïve listeners. Special care was taken to design the experiment in such a way that it would allow to quantitatively estimate the influence of the visual bias.

2 METHOD

2.1 Participants

Twenty subjects (10 male, 10 female, mean age: 24 years) of various nationalities participated in the experiment. All were audiometrically screened and had a pure-tone hearing threshold of less than 20 dB HL in the frequency range from 250 Hz to 8 kHz.

2.2 Automotive sounds

A total of 22 interior automotive sounds were selected for the experiment. Fourteen of these were original recordings of different passenger cars accelerating full throttle in 2nd gear. They were taken from a commercial data base (AVL - Institute for Internal Combustion Engines, Graz, Austria) and are based on a recording technique employing a highly simplified artificial head (AVL SOURCE) placed in the passengers seat. Two of the 14 sounds were manipulated in level using gain factors ranging from -2 to + 2 dB in 1-dB steps, thus creating 8 additional stimuli. All sounds were faded in and out with rise/decay times of 50 ms. The energy-equivalent A-weighted sound-pressure levels of the 22 sounds ranged from 64,2 to 77,5 dB, and their durations varied (due to differences in car performance) between 6,6 and 17,2 s.

The sounds were D/A-converted by a RME Hammerfall DSP soundcard (HDSP9632) with 16-bit resolution, and a 48-kHz sampling rate. Subsequently they were fed into a headphone amplifier (T.C. Electronic Finalizer Type VIZ001), and delivered to the subjects sitting in a sound-attenuated chamber via Sennheiser HE60/HEV70 electrostatic headphones.

2.3 Visual stimuli

A total of 44 color images of cars of different makes and models were collected from the Internet. They were exterior photos of passenger cars all facing towards the left and mostly shot 'in motion' in a natural environment. The images were presented on a 15-inch LCD flatscreen, and the image resolution was kept constant at 1024x768 pixels.

2.4 Experimental design

In order to assess visual bias, each of the 22 automotive sounds was paired with two pictures, one displaying a high-performance car (positive bias), and one displaying a more regular model (negative bias). To increase generality, two 'setups', i.e. two different assignments of the positively and negatively biasing pictures to the 22 sounds were produced, thus generating 88 sound-picture combinations in total. The particular combinations were chosen with the caveat that – intuitively – no highly incredible pairings of sound and picture resulted. In addition to the combined ratings, there were ratings of the 22 sounds alone, and of the 44 pictures presented in isolation. The combined ratings were always collected first, followed by either the sound, or picture ratings, counterbalancing the order of the latter two across subjects.

2.5 Procedure

Prior to the experiment proper, the participants were informed that the experiment was about automotive sound quality. Initially, they were introduced to the combined ratings only by requiring them to look at the picture, listen to the sound and decide: "How powerful does this car sound?". They were instructed to make their assessments on a scale displayed on the screen right after the sound-picture presentation. The scale consisted of 9 horizontally arranged squares of graded shades of grey, the leftmost of which was white, and labeled "not at all," and the rightmost of which was black and labeled "extremely" (powerful).

Only after they had completed the combined ratings were the participants asked to complete additional trials on the sounds alone or the pictures alone. For the latter, the instruction read: "How powerful do you think the car on the picture is?".

3 RESULTS

3.1 Combined audio-visual ratings

The raw data were aggregated across subjects by computing arithmetic means and standard errors of the category ratings which were coded by integers 0-8. These are displayed in Fig. 1 for the combined (sound plus picture) ratings, each sound marked on the x-axis being combined with either a positively biasing (triangles), or a negatively biasing (squares) picture.

Data from the two setups are combined, yielding a total of 40 (2 x N) ratings going into each data point. It is evident that the mean ratings span a large portion of the scale, ranging from 2.9 (for the sound of a diesel sedan) to 6.47 (for that of a sports car). As is evident in Fig. 1, the effect of the biasing pictures amounts to a little less than one category on the 9-point scale.

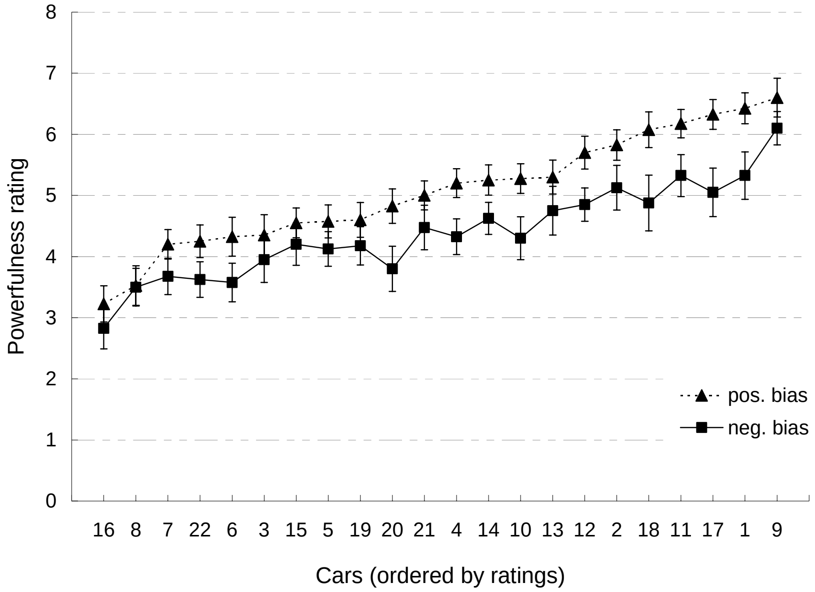

Fig. 1. Effect of positively (triangles) and negatively (squares) biasing pictures on ratings of automotive powerfulness. Each data point is based on a two ratings (of different sound-picture combinations) by 20 listeners. Averages are depicted, and the associated standard errors of the means reflect interindividual variability only.

The significance of that effect was confirmed by a 4-factor, mixed analysis of variance (ANOVA) with the sounds (22 recordings), the picture bias (2 kinds: positive and negative), and the setup (2 ways of combining sounds and pictures) constituting within-subjects factors, and the subject's gender (male/female) being a between-subjects factor. In addition to the main effect of sound [$F(21,378) = 35.03$; $p < 0.001$], there was a highly significant main effect of the picture bias [$F(1,18) = 14.39$; $p < 0.001$], while the main effects of the setup, and of gender were insignificant ($p > 0.6$). Since the interaction between the setup and bias factors was insignificant as well, it may be concluded that the particular way of combining sounds and pictures had no effect on the outcome of interest. The participant's gender did not appear to exert any systematic influence either, and the significant interactions that were found between gender and sound, as well as gender, sound and setup, appear to reflect 'local' differences in response to particular stimuli.

3.2 Effects of presentation level

Since two of the recordings were manipulated in level, the effects can also – for a subset of the data – be analysed for the effect of presentation level on 'powerfulness' ratings. A within-subjects, 3-factor (level, picture bias, recording) ANOVA of that subset of combined ratings shows nearly the same effect size for the bias [F(1,19) = 15.34; p = 0.001] as the analysis of all combined ratings did. That is also evident in Fig. 2, which shows the bias effect as a function of the presentation level (ranging from -2 to +2 dB re the original recording). In addition, the ratings of the sounds alone are depicted. It is evident that the effect of the level manipulation is approximately the same for sounds alone, the positively, and the negatively biased sound-picture combinations. Typically, judgments of the sounds alone fall between ratings of the two combined situations.

Fig. 2 also permits an estimate of what kind of level change is equivalent to the effect of the visual bias observed. Since the 4-dB level range plotted in Fig. 2 produced a change in mean powerfulness ratings of the sounds alone (crosses in Fig. 2) by a little over one category (1.225), and since the average bias effect had been observed to be a little less than a category (0.68, see Fig. 1), that corresponds to slightly more than half of the 4-dB range, or a 2.2-dB change.

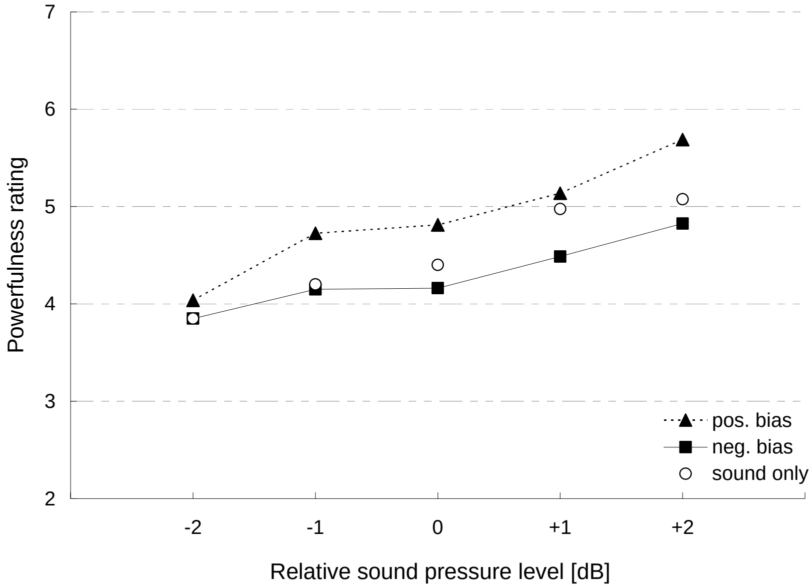

Fig. 2. Effect of presentation level on ratings of powerfulness. Two original recordings were shifted in level in the range +/- 2 dB; thus each (solid) data point contains 80 (20 listeners x 2 recordings x 2 setups) ratings of sound-plus-picture combinations which are a subset of those depicted in Fig. 1. Triangles refer to ratings of sounds combined with positively biasing, and squares to ratings of sounds combined with negatively biasing picture. In addition, mean ratings of the sounds alone (2 recordings x 20 listeners) are marked by open circles.

3.3 Variance accounted for by visual vs. auditory input

Another way of analyzing the contribution of the visual component to the sound powerfulness ratings is to inspect what portions of the total variance in the ratings are accounted for by the sound and picture factors, respectively. A conservative way to do this is to compute (classical) 'eta-squared' [5], see eq. 1,

$$\hat{\eta}^2 = \frac{SS_{Effect}}{SS_{Total}} \quad (1)$$

where SS_{Effect} is the sum of squares associated with the effect of interest, and SS_{Total} is the total sum of squares in the ANOVA. For the combined data depicted in Fig. 1, the effect of the sounds accounted for $\eta^2 = 583.46/2288.67 = 25.5\%$ of the variance, while the effect of the visual bias (the vertical displacement of the two curves in Fig. 1) accounted for $\eta^2 = 102.27/2288.67 = 4.47\%$ of the variance. That implies that the effect of the visual bias amounts to approximately one-fifth of the systematic variance seen in the combined sound-plus-picture ratings.

4 CONCLUSIONS

A laboratory experiment in which sounds were submitted to sound-quality evaluation while being accompanied by pictures showing suggestive images of the supposed sound source, provided evidence for a moderate, but significant visual bias. Note that this was the case, even though the sounds were interior recordings, and the pictures were exterior images of cars. Different measures of the effect size showed the visual bias effect (a) to be approximately one category on a 9-point scale, (b) to be equivalent to a level change by 2-3 decibels, and (c) to account for roughly 20% of the variance in the ratings of sound-picture combinations.

These results suggest that (1) the effect of the sound is dominating even in suggestive audio-visual stimulus presentations, (2) listeners integrate the visual information to some extent, even when asked to judge the auditory input alone, (3) caution should be exercised when ‘enriching’ an experimental setup with non-focal information from other sense modalities. Further work (including analyses of the ratings of the ‘pictures alone’) will show how ‘individual’ these effects are, and will attempt to extend the findings to other instances of product sound evaluation.

REFERENCES

[1] A. Kohlrausch and S. van de Par, “Audio-visual interaction in the context of multimedia applications”, In J. Blauert (Ed.), *Communication Acoustics* (pp. 109-138). Springer, Berlin, Germany, 2005.

[2] S. Viollon, C. Lavandier and C. Drake, “Influence of visual setting on sound ratings in an urban environment”, *Applied Acoust.* 63, 493-511, 2002.

[3] G. Watts, L. Chinn and N. Godfrey, “The effects of vegetation on the perception of traffic noise”, *Applied Acoust.* 56, 39-56, 1999.

[4] J. G. Beerends and F. E. de Caluwe, “The influence of video quality on perceived audio quality and vice versa”, *J. Audio. Eng. Soc.* 47(5), 355-362, 1999.

[5] C. A. Pierce, R. A. Block and H. Aguinis, “Cautionary note on reporting eta-squared values from multifactor ANOVA designs”, *Educat. Psychol. Measurement* 64(6), 916-924, 2004.