Audio-visual interactions in loudness evaluation

Hugo Fastl

AG Technische Akustik
MMK, TU München
fastl@mmk.ei.tum.de

Abstract

In psychophysical experiments, the influence of the presentation of visual stimuli on the loudness evaluation of sounds from traffic noise has been investigated. In a first experiment it was shown that the colour of the visual stimulus can influence loudness evaluation in such a way that – for same acoustic stimulus – the loudness of a red train can be rated 15 % higher than the loudness of a green train.
The following set of experiments was devided into two groups: on the one hand, visual stimuli from still pictures and on the other hand visual stimuli from moving pictures. As subgroups of the still pictures either pictures not related to the sounds or pictures fitting the sounds were presented. An example for the first subgroup would be the sound of a train passby plus a picture of a tree, and for the second subgroup the sound of a train plus the picture of a train.
As concerns the moving pictures, also two subgroups of stimuli were studied: in one case, the video was taken from a fixed position, and in the other case from a moving position such as in a driving car.
Moreoever, the presentation was either in a sound proof booth for psychoacoustic experiments via headphones and a head mounted display or in a simple car simulator. The magnitude of the influences of the rather different types of audio-visual interactions on loudness evaluation of sounds from traffic noise generally increases for more "realistic" situations: relatively large loudness reductions (8 %) are found, when moving pictures taken during a car ride are presented in a car simulator.

1. Introduction

Research on audio–visual interactions these days is very popular in the context of multimedia applications. For example the synchrony of lip movements and the correlated sounds governs the quality of applications like video telephony, video conferences, satellite TV and so forth (see e.g. [1]). With respect to source localization, usually a dominance of the visual input is assumed. A pertinent example is the ventriloquist, where the source of the speech sound (erroneously) is located at the moving lips of the puppet. However, recent results on ventriloquism [2] indicate that the concept of visual dominance has to be re-evaluated.
In the context of audio–visual interactions with respect to noise evaluation several studies have been performed using the method of semantic differential. For example Abe et al. [3] could show that when white noise is accompanied by the picture of a waterfall, the rating improves towards more positive adjectives. Viollon et al. [4] could demonstrate significant differences in the evaluation of bird twittering, if in addition to the acoustic stimuli on the one hand the picture of an urban environment and on the other hand the picture of a forest were presented. Of great importance are results of Hashimoto and Hatano [5] who showed that the presentation of a picture can reduce the perceived annoyance of car sounds. The effects were very dramatic, since a sound could be presented together with a picture at a level 10 dB higher in order to elicit the same reactions as the sound when presented without picture.
In this paper, results of audio-visual interactions are described which concentrate on variations in perceived loudness. Several situations are reported which get step by step more realistic, starting from still pictures not related to the sound source to videos with corresponding audio presented in a simple car simulator.

2. Experiments

Since a detailed description of the experiments is given in a publication by Christine Patsouras [6], in this paper only some pertinent features are mentioned. Acoustic stimuli were recorded by a half inch condenser microphone or by a dummy head and stored on DAT tape. After editing of the sounds, they where reproduced by headphones with appropriate equalizing. Care was taken to present all sounds at their original levels. Optical stimuli where recorded by a digital camera for still pictures, and by a video camera for moving pictures. Presentation of visual material was either by a head mounted display with a virtual screen of 1.2 x 1.2 meters or by a beamer included in a simple car simulator.

3. Results

3.1. Influence of colour

In a first series of experiments it was studied, if the colour of pictures presented in addition to the acoustic stimuli can influence the loudness rating to a different degree. The still picture of a German high speed train (ICE) was taken; its original colour is white with a red

From: ICA 2004, Kyoto, Intern. Commision on Acoustics

stripe. By using paintbrush, the colour of the ICE train was modified to red, blue or green. The trains with different colour are depicted in figure 1.

Figure 1: ICE trains with different colour

The original sound of the passby of an ICE train was combined with pictures of the train in different colour. The presentation of the sound was by headphones, the presentation of the still image by a head mounted display. The subjects were asked to rate the loudness of the passby by the method of magnitude estimation. Average results for two types of train sounds are illustrated in figure 2. The loudness rating is given relative to the loudness when the sound is presented in combination with the green train.

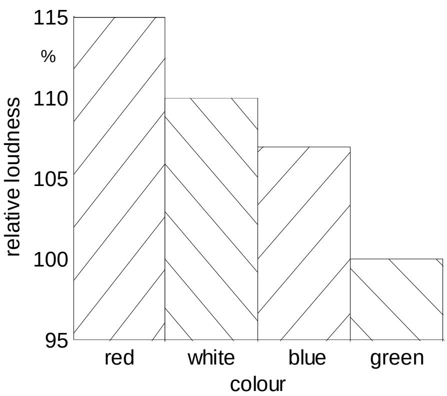

Figure 2: Perceived relative loudness for the <u>same</u> acoustic stimulus if in addition pictures of ICE trains with different colour are presented.

The results displayed in figure 2 clearly show that – despite the same acoustic stimulus – a red train is perceived as being on the average by 15 % louder than a green train. Combinations of one and the same sound with white or blue trains elicit a loudness which is somewhere in between.

The results displayed in figure 2 are quite in line with the practice of manufactures of expensive sportive cars: A large magnitude of expensive European sportive cars shows the colour red, whereas such a sportscar is not advertised in a light green colour.

3.2. Still pictures <u>not</u> related to the sound

In a next series of experiments it was studied, whether pictures which are not related to the sound can influence the loudness evaluation. Again, the passbys of different trains were chosen as sounds. They were combined with pictures, which have no direct relation to the sound source. In a first experiment, together with the train sound, the picture of a tree was presented. On the one hand, the picture was taken in summer with the tree in full green leaves. On the other hand, the picture taken in winter shows the tree with it branches without leaves in a landscape of snow.

Figure 3: Pictures of a tree in summer and winter as examples for pictures not related to the sound source, in this case an ICE train.

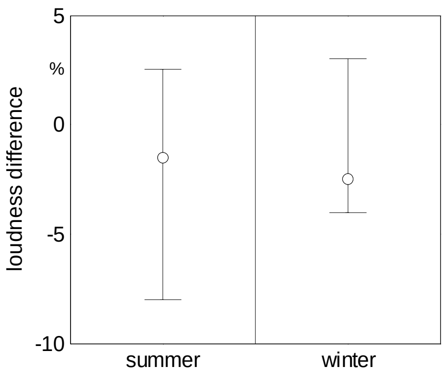

Figure 4: Influence of still pictures showing a tree in summer or winter on the perceived loudness of a train passby.

Results plotted in figure 4 show the difference in perceived loudness if in addition to the sound the picture of a tree in summer or in winter is presented. Data are given as medians and interquartile ranges. For identical acoustic stimuli, the presentation of an unrelated picture can reduce the perceived loudness on the average (median) by 1.5…2.5 %, and there is large overlap of the interquartiles. The effects of the tree in summer, where the lower quartile reaches values of - 8 % could be related to the green colour. On the other hand, the median shows a larger influence of the picture for the tree in winter. Perhaps several subjects inferred that snow is an absorber of sound.
The next experiment assessed the influence of the colour green further. Together with the sound of a train passby, either the picture of a street lined by trees without leaves, or of the same street where leaves have been electronically added to the trees, was presented.
According to the data plotted in figure 6, the presentation of the still pictures reduces the perceived loudness of the train passby on the average (median) by 2.5 %. Taking into account in addition to the medians the interquartiles it becomes clear that again the green colour of the leaves seems to reduce the perceived loudness even more.

Figure 5: Still picture of a street lined by trees without or with electronically added green leaves not related to the sound of a train.

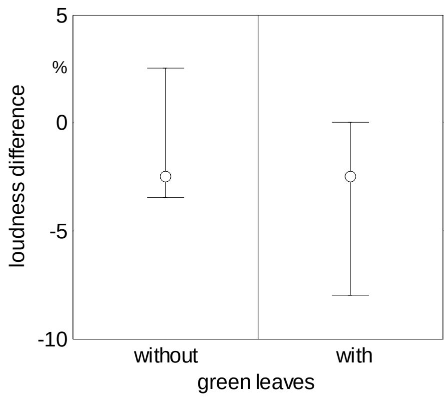

Figure 6: Influence of still pictures showing a street lined by trees without or with electronically added green leaves on the perceived loudness of a train passby.

In a last experiment with unrelated pictures, a possible influence of pictures from different regions was studied. The pictures used together with the passby noise of trains are displayed in figure 7. On one picture, old farmhouses indicate a rural region. The other picture shows a metropolitan region with high rise buildings and a TV tower. The third picture was taken in an industrial region.

Figure 7: Still pictures not related to the sound of a train passby, taken in a rural, a metropolitan, and an industrial region, respectively.

Results displayed in figure 8 show that the unrelated pictures from different regions only slightly influence the perceived loudness of a train passby. The medians of the loudness differences range between 0 and -2.5 %. There is strong overlap of the interquartile ranges. From all three pictures considered, on the average (median) the rural region shows the largest influence on perceived loudness, perhaps since many persons expect from rural regions more quietness than from metropolitan or industrial regions.

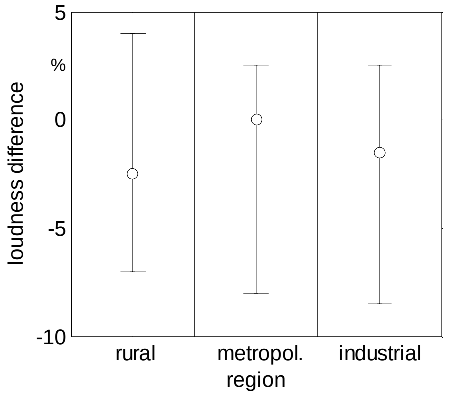

Figure 8: Influence of still pictures showing a rural, metropolitan, and industrial region, respectively, on the perceived loudness of a train passby.

3.3. Influence of still pictures related to the sound

In a next series of experiments, together with the sound of a passing train, a still picture of a train was presented via a head mounted display. An example of the pictures used is given in figure 9.

Figure 9: Example for a still picture related to the sound, i.e. the passby of a train.

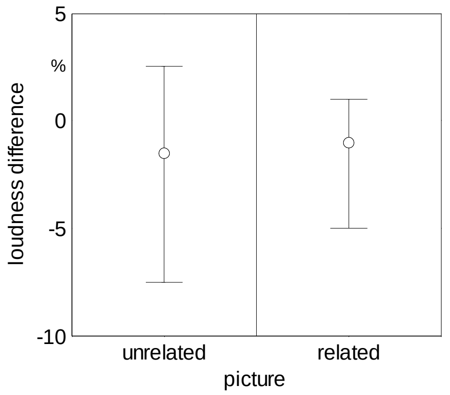

Figure 10: Influence of unrelated vs. related still pictures on the perceived loudness of a train passby.

Figure 10 shows influences on the perceived loudness of a passing train by unrelated vs. related pictures. For the unrelated pictures, an average across many different situations was taken. For the related picture, always images of trains were used, however for different types of trains like passenger trains, freight trains and so forth. The results displayed in figure 10 suggest that on the average both related or unrelated pictures decrease the perceived loudness of train passbys only slightly. There is strong overlap of the interquartiles. The somewhat larger interquartiles for the unrelated pictures may be partly due to the fact that many rather different situations were compiled.

3.4. Still versus moving pictures of related sounds

Figure 11 shows as an example a snapshot of a video showing a commuter train moving from left to right. The video camera was turned in such a way that the whole passby was captured.

Figure 11: Example for a snapshot from a video displaying a comuter train, i.e. a moving picture related to the sound

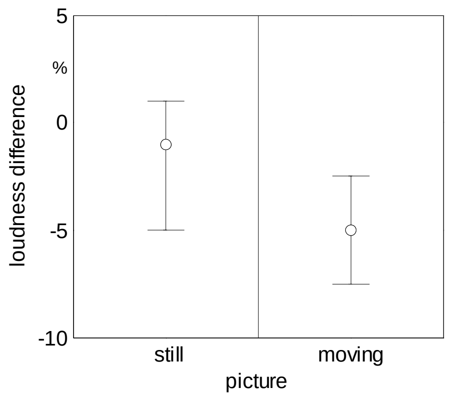

Figure 12: Influence of still vs. moving pictures of trains on the perceived loudness of a train passby.

The results illustrated in figure 12 allow a comparison of the impact of still vs. moving pictures of a train on the perceived loudness for physically identical passby sounds.

Results displayed in figure 12 show that the presentation of a moving picture together with the audio signal clearly reduces the perceived loudness, in this case by 5 %. On the contrary, a still picture of a train presented in addition to the acoustic stimulus reduces the perceived loudness on the average only by 1%. Obviously, the moving picture together with the dummy head recording induces a more realistic situation in which larger audio-visual effects occur. The larger loudness reduction may point to a stronger impact of the visual input.

These effects are well known by sound recording engineers, who realize the sound tracks of movies: When heard without watching the moving picture, the sounds are unrealistically loud. However, because of the audio-visual interaction between the audio signal and the moving picture, the perceived loudness is reduced. Therefore, in the context of the movie, the loudness of the sounds is just right.

3.5. Influence of fixed vs. moving position of video camera

In this series of experiments, the influence of the position of the recording equipment was studied. On the one hand, both the dummy head and the video camera were at a fixed position anchored on the ground. In the other case, the dummy head was placed on the front passenger seat of a car, and a video was taken through the windscreen while driving the car. In this case, the position of the recording equipment for both audio and video was moving.

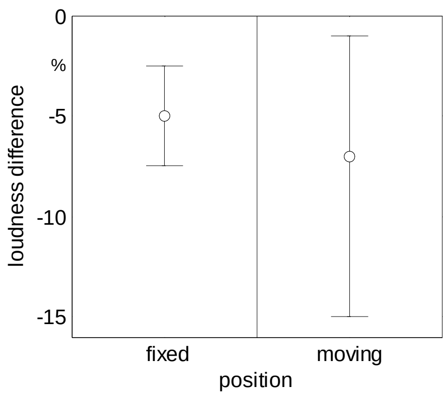

Figure 13: Influence on loudness evaluation when the sound and the moving picture is either recorded by equipment fixed on the ground or by equipment moving in a car.

Results displayed in figure 13 enable a comparison of the impact of moving pictures on the evaluation of perceived loudness when the recording is done from a fixed vs. a moving position.

Results displayed in figure 13 indicate that the reduction in loudness perception is larger, if the moving picture is taken from a moving position like the passenger seat of a car in comparison to a fixed position on the ground. Not only the median loudness reduction increases from –5 to –7 % but also large interquartiles down to –15 % indicate that for some persons substantial loudness reductions can be obtained.

3.6. Presentation of stimuli in a booth versus a simple car simulator

By means of figure 14, the presentation of sounds in a simple car simulator is illustrated [7]. The test subject sits in the drivers seat of a real car (BMW 7 series), and the sounds are presented via headphones. In addition, a video is projected by a beamer on a screen in front of the vehicle. In such a setup, the subject can get a feeling similar to really driving in a car.

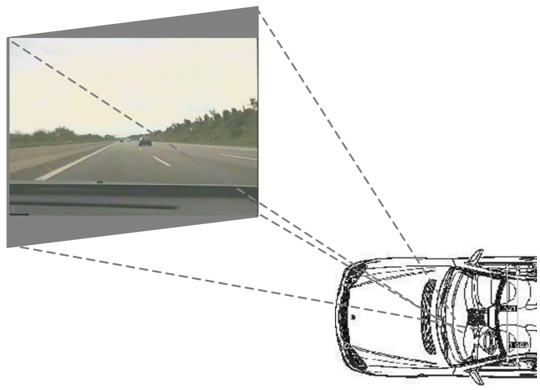

Figure 14: Illustration of the setup for the simple car simulator.

Figure 15 shows the loudness reduction induced by the presentation of a video by a head mounted display when the subject is seated in a sound proof booth vs. the situation that the subject sits in a simple car simulator.

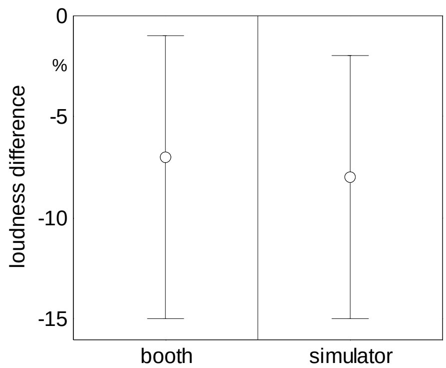

Figure 15: Reduction of perceived loudness when watching a video taken during driving in a car either by a head mounted display in a sound proof booth or by presentation in a simple car simulator.

The data displayed in figure 15 suggest that the presentation of videos taken from a moving position can substantialy reduce the perceived loudness compared to the presentation of dummy head recordings without visual input. Both situations, i.e. presentation of the video via a head mounted display in a sound proof booth vs. presentation in a simple car simulator, lead to similar results. On the average (median) the perceived loudness reduction is somewhat larger in the car simulator than in the measurement booth. Perhaps the typical feeling of sitting in a car enhances the audio-visual interactions since it is closer to the real live situation.

4. Discussion

The results displayed in section 3 clearly show that – despite identical acoustical stimulus – the addition of visual inputs can influence the perceived loudness of sounds. Usually the additional visual input reduces the perceived loudness. As concerns the effect of different colours, objects in green colour seem to be softer in particular when compared to the same objects in red colour. While still pictures can reduce the perceived loudness on the average by about 2.5 %, moving pictures can induce reductions in perceived loudness around 5 %. These figures hold for the condition that the recording equipment both for the acoustic and the optic stimuli is in a fixed position. If on the other hand the recording equipment for audio and video is in a moving position, like a dummy head and a video camera in a car, even larger loudness reductions up to an average of about 8 % can be obtained. By and large it can be stated that the more realistic the situation, the larger the possible loudness reduction induced by visual images for the same acoustic stimuli.

5. Outlook

While in this paper we concentrate on audio-visual interactions which manifest themselves in differences of perceived loudness, audio-visual effects may have a much broader impact. As described by Kuwano et al. [8], not only loudness but also rating of aesthetic features can be significantly influenced by audio-visual interactions. Like in our studies on loudness evaluation, also in the study by Kuwano et al. on aesthetics, green leaves of trees play an important role.

In the cases mentioned so far, additional visual stimuli altered the perception of acoustic stimuli. However, the contrary may also happen. As an example, we mention a visual illusion induced by sound [9]: If in addition to the presentation of a single flash of light pulsed acoustic stimuli are presented, the visual image is also perceived as pulsating.

An even larger dominance of acoustic perception over visual perception is present in synaesthetics (e.g. [10]). In persons who show this phenomenon, acoustic stimuli can elicit visual sensations. For example, different pitches are correlated to different colours, i.e. tones played on a piano may produce not only the acoustic sensation of pitches of different height, but the person also sees different related colours! In such a case, the acoustic perception clearly dominates and governs the visual perception..

.

In summary then, audio–visual interactions may have a strong impact on the evaluation of products as well as environmental situations. Although audio-visual interactions have been studied extensively in the context of multimedia applications, much work remains to be done in the field of applied psychoacoustics.

6. Acknowledgements

The author wishes to thank Dr.-Ing. Christine Patsouras for systematic studies on audio-visual interactions as well as stimulating discussions on this topic. Dipl.-Ing. Markus Fruhmann and Dipl.-Ing. Stefan Kerber are acknowledged for substantial editorial assistance. This work was supported by Deutsche Forschungsgemeinschaft, FA 140/2.

7. References

[1] Kohlrausch, A., Begault, D., "Workshop on audio-visual interactions", 110th AES Convention Amsterdam, 2001
[2] Seeber, B., "Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode", Dissertation, TU München, 2003
[3] Abe, K., Ozawa, K., Suzuki, Y., Sone, T., "The effects of visual information on the impression of environmental sounds", Inter-noise 99: 1177-1182, 1999
[4] Viollon, S., Lavandier, C., Duval, R., "Development of an experimental procedure suitable to test audio-visual interactions in the complex urban sound environments", Euro-noise 98, Vol. 2: 1095-1100, 1998
[5] Hashimoto, T., Hatano, S., "Effects of factors other than sound to the perception of sound quality", 17th ICA Rome, CD-ROM, 2001
[6] Patsouras, Ch., "Geräuschqualität von Fahrzeugen – Beurteilung, Gestaltung und multimodale Einflüsse", Shaker Verlag Aachen, 2003
[7] Patsouras, Ch., Böhm, M., Filippou, T. G., Fastl, H., Patsouras, D., „Einfluß einer optischen Komponente bei der Beurteilung der Lautheit von Fahrzeuginnengeräuschen bei Fahrt", Fortschritte der Akustik, DAGA 03, DEGA Oldenburg, 2003
[8] S. Kuwano, S. Namba, M. Komatsu, T. Kato and Y. Hayashi, "Auditory and visual interaction in the aesthetic evaluation of environment", Empirical Studies of the Arts, 19(2):191-200, 2001
[9] Shams, L., Kamitani, Y., Shimojo, S., "Visual illusion induced by sound", Cognitive Brain Research 14: 147-152, 2002
[10] Haferkamp, M., „Visualisierung auditiver Wahrnehmung – historische und neue Konzepte", Fortschritte der Akustik, DAGA 03, DEGA Oldenburg, 2003