Perception of emotion portrayal in cartoons by aurally and visually oriented people

PerMagnus Lindborg

Nanyang Technological University, Singapore

ABSTRACT

This article reports results from a study of perceived emotion portrayal in cartoons by different groups of subjects. A set of audiovisual stimuli was selected through a procedure in two steps. First, 6 'judges' evaluated a large number of random snippets from all Mickey Mouse cartoons released between 1928 and -35. Analysis singled out the five films ranking highest in portraying respectively anger, sadness, fear, joy and love/tenderness. Subsequently, 4 judges made a continuous evaluation of emotion portrayal in these films, and six maximally unambiguous sequences were identified in each. The stimuli were presented to two groups (N=33), one in which the subjects were expected to be visually acute, and one where they would tend to be more aurally acute, in three different ways: bimodally (original) and unimodally, i.e as an isolated sound or video track. We investigated how group and modus conditions influenced the subjects' perception of the relative intensity of the five emotions, as well as the sense of realism portrayed in the cartoon clips, and how amusing they were found to be. Finally, we developed an estimate for *visual-aural orientation* as a linear combination of select self-reported variables, and tested it as a predictor for the perception of medium dominance.

1. BACKGROUND AND AIMS

Before embarking on the present study, we had decided to investigate emotion perception of audiovisual artworks where the two media were created to function together as a whole. We chose to work with early cartoons for two main reasons. First, the creators at Walt Disney Studios, pioneering a brand new bimodal medium, consciously strived to install a balance between aural and visual domains. Second, while the huge influence of Mickey Mouse on animation art is well known, the very earliest works are not widely known to people today. When the 'talkies' appeared, Disney saw an opportunity to "bring a symphony orchestra to every small town in America". His studio's works were promoted as "sound cartoons", intended to be both heard and seen. Early animators were very conscious about audio-visual integration, the creative process "intertwining sound and visuals" (Thomas & Johnston 1981). Carl Stalling, the composer behind the first dozen Mickey Mouse cartoons, described how "sometimes the director made the action fit a certain piece of music, and at other time I wrote music to fit certain actions." (Barrier 2002). In the 'sound cartoons', composers could indeed experiment with a rich "sonic fabric that includes the musical score, ambient sound, dialogue, sound effects, and silence" (Lipscombe & Tolchinsky 2005).

In 1976, McGurk & MacDonald described multimodal interaction in cognitive terms, in particular how the dominant visual sense 'spills' information into the aural domain. What we see may distort what we hear. The Congruence-Associationist framework (Bolivar, Cohen & Fentress 1994, Cohen 2001) is the most developed model, attempting to explain how film meaning is created through the perception of certain cinematic components, labelled "speech", "visual narrative" and "music". However, Lipscomb & Tolchinsky expressed reservation about the model's fundamental assumption of visual primacy, lending to the auditory component a mere subservient role, and called for additional supporting research. A "special relationship", dubbed *synchresis* by Michel Chion, is installed in our mind when sounds and images occur at the same time. The channels of sensorial input fuse, and are "perceived as having deep ontological kinship." Hence, "the disarticulation of sound and images leads… to a sense of absurdity." (Corbett 2002). According to Annabel Cohen, one of the functions of film music is to heighten the "sense of reality of or absorption in film, perhaps by augmenting arousal." (Cohen 2001). This suggests a testable proposition: how does the sense of *realism* change in cartoons under different conditions, with and without the sound track?

In much narrative cinema, music does not draw attention to itself, and when it does, it is often in order to support extreme narrative moments such as great absurdity, passion or violence. The way the audience is affected must be understood in terms of emotion perception. In a review of a large number of studies on emotion communication in vocal expression and music, (Juslin & Laukka 2003) argued that the best perspective is provided by evolutionary psychology, an approach that explains why humans feel in certain ways as adaptive reactions to basic survival problems. The authors emphasised the advantage of using a limited number of qualitative emotion categories in research models. For our study, we employed the five basic emotion labels recommended in the review: *anger, sadness, fear, joy* and *love/tenderness*. The impact of music in film has been analysed by several authors. By contrast, studies of bimodality in film have rarely treated soundscape and effects on par with dialogue and music (Biancorosso 2009). Part of the reasons for this may lie in inconsistency of terminology (which is the more overarching, 'sound' or 'music'?) and the complexity of audiovisual cognition. In comparison with films, cartoons have not received the same music research attention. Cartoon

soundtracks are quite different from feature films. Most notably, cuts are faster and more drastic, an observation that induced John Corbett to state a "precise formulation for the cartoon aesthetic: suddenness" and make a comparison with Stockhausen's notion of *Momentform* (Corbett 2002). But how is emotion perceived to be portrayed in cartoons?

2. HYPOTHESIS

We investigated the hypotheses that 1) people perceive emotion in isolated sound and video tracks differently from what they perceive when both media are present, and that 2) this discrepancy is emphasised by individual profile expressed as aural-visual orientation or pre-disposition. Further, we surmised that a) people profiled as visually oriented will perceive emotions more intensively when presented with visual stimuli, and opposite so for the aurally oriented, and that b) the visually oriented would have a stronger agreement in emotion judgement between visual-only and bimodal stimuli, and vice versa.

3. METHOD

For the Cartoon Emotion Experiment (CEx), we created a set of audiovisual stimuli through a procedure of pre-tests. We decided to use early Mickey Mouse cartoons, created between 1928 and 1935 at Walt Disney Studios. The 34 films have been released as (*Mickey Mouse in black and white: the classic collection* 2002). For our research purposes, the films were copied to hard disk using SnapzPro to QuickTime movie format with H.264 compression and resolution not lower than 920x690 pixels in 20 fps, and audio in stereo 16 bits linear PCM. The total running time of the cartoons being more than 4 hours, the first pre-test was designed to narrow down the corpus.

3.1 Pre-test 1

The first pre-test was designed to single out the films that scored highest in portraying each of the 5 basic emotions (anger, sadness, fear, joy, love/tenderness). 6 professional media artists and musicians (4 men, 2 women) volunteered to act as 'judges'. The films were screened on an LCD display (Apple 21" 1900x1200 pixels) and the sound over headphones (Sennheiser HD650). The software, programmed in MaxMSPJitter, was designed to start by randomly picking one of the 34 films and a time location to start screening at (always bimodally, both sound and video). The judge had to rapidly decide which emotion s/he perceived, and press a corresponding keyboard button (*A, S, F, J, L*). The film would then gain a point in that emotion category, before a new clip was generated and screened. The process selecting the next films continuously kept track of the 'expected' emotion for each film. The random selection was weighted by the 'hit' score received in a category up to that point. The expected emotions were made to rotate randomly in groups of five ("urn" procedure) to avoid pattern bias. If the judge's evaluation of emotion portrayal was what the system predicted, that film's weight score received a 'bonus'; if it did not, a 'penalty'. The film picking algorithm uses a probability weights vector $W = \{w1, w2\ldots,w34\}$ which was updated as follows for the two cases:

judged = expected: $w_{judged} \leftarrow w_{judged}(1+a)+b$;

judged ≠ expected: $w_{expected} \leftarrow w_{expected}(1-a)$.

After heuristic trials before the test, the constants were set at a=0.22 and b=1.

Results

Each judge evaluated around 200 clips over approximately 30 minutes. The 34 films received an average of 31 'hits' each. The mean time spent to decide emotion portrayal was 9.8 seconds (SD = 3.0). This led us to decide that a clip duration in the order of 10 seconds was going to be appropriate. Because the Mickey Mouse cartoons are generally lively, clips of longer duration may contain conflicting content. As we needed to keep the duration of the experiment within roughly one hour for practical reasons, the main test should maximally use 30 clips. The films, selected according to a simple ranking algorithm, are listed in Table 1.

film	Judged emotion	strength
Mickey's Service Station (1935a)	*anger*	0.158
The Chain Gang (1930b)	*sadness*	0.178
Firefighters (1930a)	fear	0.144
Mickey's Orphans (1931e)	joy	0.156
The Birthday Party (1931a)	Love/tenderness	0.156

Table 1: Selected films

3.2 Pre-test 2

The second pre-test was designed to assist the selection of short excerpts based on a criterium of 'emotion unambiguity'. 4 judges (3 men, 1 woman) volunteered. The software was made in MaxMSPJitter, and the setting similar to that used in the first pre-test. Each film was screened from beginning to end. The judge was instructed to continuously evaluate emotion portrayal in the five categories, and press a corresponding button when it changed. The result was a timeline graph with the emotions as levels. There was strong correlation with the emotion labels as judged by the first pre-test, indicating that the film could proxy for a emotion. The graphs of all judges were added together to produce a smooth intensity curve for each film's main emotion.

Results

Visual inspection of the graphs revealed high plateaus and peaks corresponding to relatively clearly defined emotion portrayal and from this, 6 sequences of 10 seconds duration were identified. In some cases the start and end points were slightly adjusted to avoid illogical scene changes.

3.3 Cartoon Emotion Experiment, *CEx*

Subjects

The main test involved 33 subjects: 17 females, 16 males (mean age=22.2 years, SD=2.1). Each person volunteered and received a token reimbursement of 8 SGD (4 EUR). The first test round called for students at a university school for art, design and media, and the 27 students who volunteered were mainly undergraduates in animation, film and interactive media. 6 responses had to be discarded due to a software error. The second round called for university students having at least three years of experience playing an instrument or singing, and being currently active in chamber orchestra, a band or similar setting. 12 students from various colleges answered the call. The first group, "adm" for short, was expected to self-report a more visually oriented disposition, while subjects in the second, "mus", were expected to tend towards an aural orientation.

Stimuli

The 30 clips, 6 from each of the 5 films, had been selected to portray five basic emotions (anger, sadness, fear, joy, love/tenderness) in a maximally unambiguous way. They were presented in three modes: *sound* only, *video* only and *both* (sound and video together). The clips had an effective duration of 10 seconds, plus an additional 0.5 s of fade-in and 0.5 s of fade-out.

Apparatus

The test software was programmed in MaxMSPJitter. Subjects were seated in a computer lab setting in front of LCD displays (Apple 21", 1900x1200 pixels) screen and wearing headphones (Sennheiser HD215) with the sound level individually adjusted to a comfortable level. Response data were entered with a mouse, by clicking on menu items and by moving sliders.

Procedure

An introduction to the test and all instructions were contained in the software. The test was in three parts. First, subjects were asked to self-report quasi-objective data of three kinds:

- *knowledge* ("How much of a 'cartoon specialist' would you say you are?")
- *activities* ("On average, how many hours a day do you spend on…[TV/games, painting, music-making, sports, working, socialising, sleep/rest]")
- *senses* ("How actively do you normally use the five senses in your work/studies?")

Then, the 90 stimuli were screened in random order, different for each subject, in order to reduce bias due to each clip being screened three times (in different modus conditions). After each stimulus, four panels with questions related to emotion portrayal perception followed in random order. Three were constant:

- *emotion* ("Which word(s) would you use to describe the emotion(s) you found portrayed in the excerpt? how strongly were they expressed?") [anger, sadness, fear, joy, love/tenderness]
- *amusement* ("How exciting or boring did you find this excerpt?")
- *realism* ("How life-like (naturalistic) or exaggerated (abstract) did you find the excerpt?")

Responses to *emotion* would indicate which words subjects use to describe what they see or hear. The *amusement* question was aimed at giving an estimate of the subject's arousal, i.e. to reflect her/his inner state, and the *realism* question at measuring valence, i.e. a more distanced evaluation of the relationship between subject and stimulus.

The fourth question was modus dependent and had different panels for the *sound*, *video* and *both* conditions. Care was taken to present the unimodal question with panels of identical layout and scope, e.g. the number of elements was 5 in both of them.

- *soundelements* ("How do you perceive the importance of different elements in the soundscape? [music, sound effects, dialogue, singing and other vocals, voice-over]")
- *videoelements* ("How do you perceive the importance of different visual elements in the excerpt? [backgrounds, camera movement, characters, objects, text & symbols]")
- *dominant* ("Which medium attracted most of your attention in the excerpt? by how much?")

It took approximately one hour to complete the 90 stimuli. A progress bar in the software interface gave feedback and attempted to steer subjects on a steady pace, pointing out if the responses were too fast or slow. After the stimuli had been completed, *subject data* (age, gender, handedness, race, home language) were collected. Finally, one group of questions was aimed at picking up *feedback* on the experimental design.

The test design thus brought together 24 independent variables on each subject, 3 variables defining 90 stimuli, and 19 dependent variables from question responses. Complementing the self-reported (consciously provided) data, the software recorded the *time* a subject spent on each occurrence of the 11 different panels, providing a set of (objective) dependent variables.

4. RESULTS

4.1 Emotion

There was good agreement between the emotion labels given by the pre-test 'judges' and the evaluations made by the CEx subjects. It should be noted that the judges could indicate one emotion at a time, but the CEx subjects could move all 5 sliders for each clip to indicate the relative strength of the emotions they perceived. Visual inspection of boxplots (a chart with 150 boxplots of means of 5 emotions in 30 clips) revealed that Film1 was the most unambiguous, with 5 out of 6 clips clearly portraying *anger*. Film2 had 2 clips very clearly portraying *sadness*, but the remaining 4 confused with *fear* and in one clip *joy* was stronger. Film3 very clearly portrayed *fear* in all but one clip, but was also generally confused to a lesser degree by *anger* and *sadness*. Film4 was the least clearly defined; labelled "joy" by the judges, the CEx subjects did find a lot of *joy* in it, but they reported equally much of *love/tenderness*, and in some cases *sadness* or even *anger*. Film5 portrayed *love/tenderness* in all but one clip, but was generally confused with *joy* to a lesser degree. Overall, it seems that the labels "joy" and "love/tenderness" were often used interchangeably, and likewise but to a somewhat lesser degree were "sadness" and "fear". The consensus was broadest for when to use the label "anger".

Looking at how the test conditions influenced *emotion* perception, we ran a MANOVA which revealed significant *group* effects on *anger, sadness* and *love/tenderness* (and almost so on *fear*). This corresponded well with the observations of the boxplots earlier. There were significant *modus* effects on *fear, joy* and *lovetender*. The results are given in Table 2.

	group (F, p)	*modus* (F, p)
anger	7.55, p<0.006**	1.87, p<0.155
sadness	12.8, p<0.0004***	2.18, p<0.113
fear	2.95, p<0.086.	3.22, p<0.040*
joy	0.007, p<0.93	3.73, p<0.024*
love/tenderness	3.92, p<0.048*	19.8, p<3e-9***

Table 2: Correlations between motions and conditions.
`(anger,sadness,fear,joy,lovetender) ~ group+modus`

This type of bimodal consistency was discussed by (Cook 1998) Our results agree to some extent, but hint that consistency may not be equally strong for different kinds of emotions.

4.2 Realism and amusement

Realism

Across all subjects, the perceived degree of *realism* had a clearly bipolar distribution, with 'negative peak' and 'positive peak' approximately equidistant from zero (means=-0.45 and 0.42, SD=0.23 and 0.22). *Realism* separates the films in two groups: films significantly portraying *anger*, *fear* and *sadness* on the one hand, and *joy* and *love/tenderness* on the other (as tested for Tukey's Honest Significant Difference with 95% family-wise confidence intervals, adjusted p<0.0014). A division along these lines corresponds to the generally accepted interpretation of emotions of negative and positive valence. As can be seen in Figure 1, the sense of *realism* was generally weakest in the *sound* condition. Less expected was the result that in four out of the five films, the condition with *video* on its own enticed a stronger *realism* than when both were present. This might indicate that for the Mickey Mouse cartoons, an important contribution of the sound track is to make the film more abstract: in short, a *Verfremdungseffekt*.

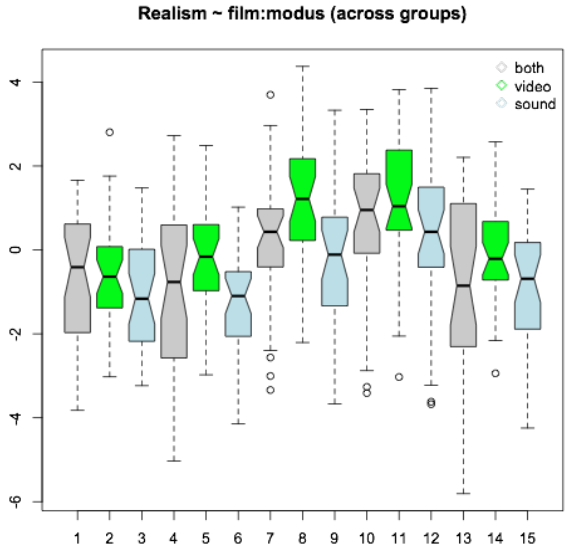

Figure 1. Boxplot of *realism* means (all subjects confounded). Numbers on the abscissa refer to the 5 films and the 3 conditions, e.g. "4, 5, 6" indicate the distributions in *both, sound* and *video* conditions from all 6 clips from the second film ("fear"). The scale on the *y*-axis is [-6..+6] because each clip was evaluated on a scale [-1..+1] and all 6 clips are summed for each condition.

Amusement

Compared to *realism,* the distribution of *amusement* was closer to normal (mean=0.199, SD=0.385) though with a negative skew. No separation of films was found using Tukey's HSD. As shown in Figure 2, subjects found the *both* conditions the most exciting for all films, which was expected. With the exception of Film2 ("sadness"), they found the *sound* condition to be more boring than *video* only. For the exceptional case, there was no significant difference between the two groups. The *amusement* was highest in Film1 ("anger") and Film2 ("sadness").

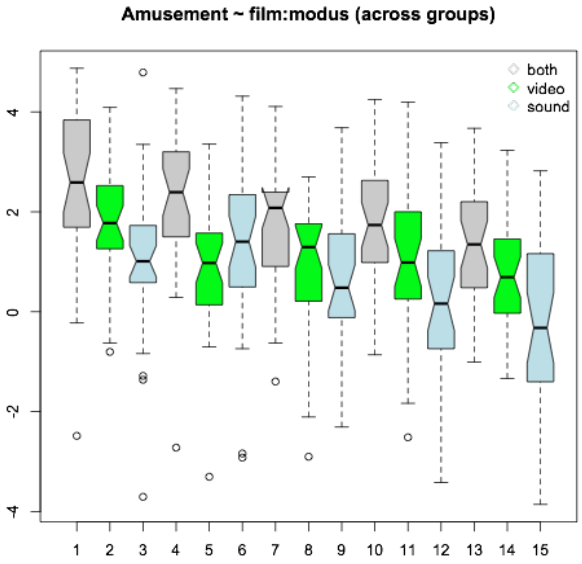

Figure 2. Boxplot of *amusement* means.

Interaction

Despite the bipolarity in the distribution of realism, there was a significant linear correlation between the two measures ($F(1, 2968)=5.58$, $p<0.019$). Figure 3 shows a scatterplot of all points (90 * 33 = 2970), but with two separate regression lines: one for the negative valence peak, one for the positive. The dotted line is the amusement mean=0.18. The cross-like cluttering of values near zero is caused by the software interface used as a platform. Subjects responded by moving sliders, but these reported a default value even though the subject may in many cases not have considered any response at all. (The default was somewhat arbitrarily set at 0.05. The R-friendly NA would have been better.)

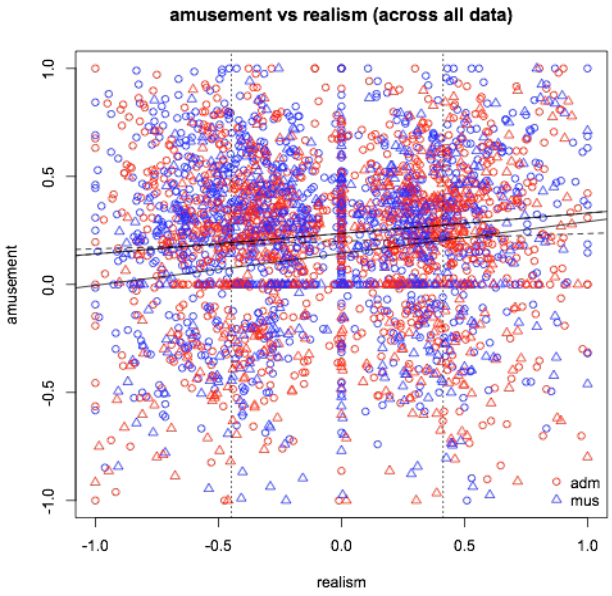

Figure 3. Plot of *amusement* against *realism*.

4.3 Modus dependence

It was rather a surprise to find that over the 990 responses (33 subjects, 30 stimuli in *both* condition), *sound* was perceived as the *dominant* medium more often than *video* (respectively 523 and 348 times), as shown in Figure 4. The group differences are clearly significant but it is unclear how to interpret them. The differences are possibly connected to results from the *feedback* question panel (asking which test topic the subject thought was most difficult to respond to), which also showed a clear difference between the groups: *adm* subjects largely indicating "sound", and *mus* "video", as indicated in Figure 5. Because *feedback* was given at the end of the test, after the subject had finished all the stimuli, we may speculate that the subjects expressed a bias against the "opposite" medium, possibly as an effect of the test soliciting them to focus on and think about "things they don't normally do". The bias would be the same regardless of the experience being perceived as positively stimulating, or just difficult in a bothering way.

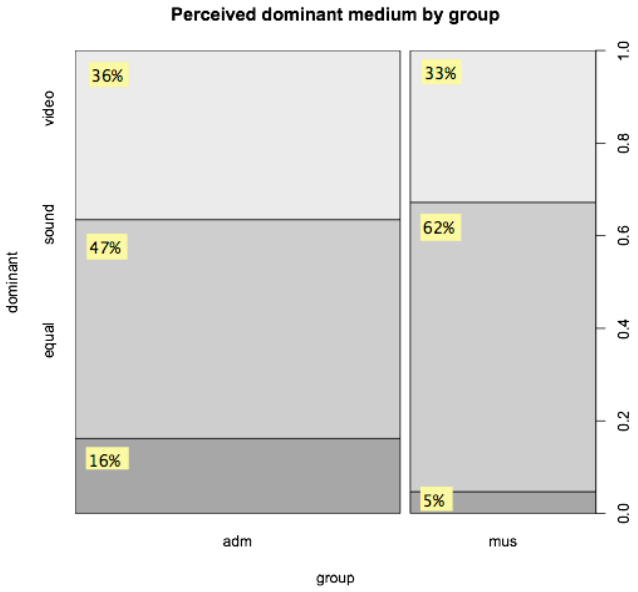

Figure 4. Sound was perceived to be the dominant medium most of the time, in particular by the aurally acute group.

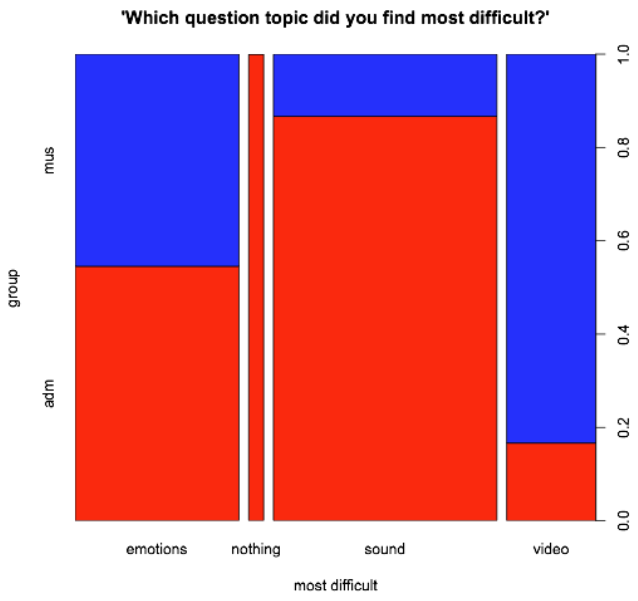

Figure 5. Topics related to sound were identified as most difficult by the visually acute group, and *video* by the *mus* group.

4.4 Aural-visual orientation

Interactions between groups and senses

We investigated the fit between *group, modus* and *senses* on one hand, and *amusement* and *realism* on the other using MANOVA. It turned out that *hearing* was not correlated with *amusement* and that there was no difference between the *adm* and *mus* groups in this regard. This might indicate that soundtracks on their own were neither heard as exciting nor boring, and that people need visuals to identify Mickey's 'gags' as funny. By contrast, for *realism, hearing* was important, especially for the *mus* group. The interaction between *group* and *sight* was significant in both cases, but somewhat surprisingly, *touch* was more strongly correlated. Here too, there was a clear difference between the groups. Table 3 presents the results from a simplified model involving three of the senses.

	amusement (F, p)	*realism* (F, p)
group	46.4, p<1.2e-11***	21.1, p<4.6e-6***
sight	69.3 p<2.2e-16 ***	35.8, p<2.4e-9***
hearing	0.27, p<0.60	27.7, p<1.5e-7***
touch	14.8, p<0.00013***	0.034, p<0.85
group:sight	4.09, p<0.011*	6.52, p<0.011*
group:hearing	0.90, p<0.34	8.69, p<0.0032**
group:touch	44.1, p<3.8e-11***	21.5, p<3.72e-6***

Table 3: Correlations between emotion and senses.
`(amusement,realism) ~ group*(sight+hearing+touch)`

To test the second hypothesis, we needed to set up a tentative formula for estimating aural-visual orientation as a linear combination of subjects' self-reported data. After the data had been collected, we investigated to find the most salient measures and reasonable weightings. The initial assumption had been that the senses *sight* and *hearing* would be important contributors to a good estimate, but inspection of the multiple linear regression between *group* and the five *senses* revealed that *sight* had negligible influence (Pearson's r=0.03). The largest regression coefficients were for *hearing* and *touch* (r=0.44 and 0.43) followed by those for *taste* and *smell* (r=0.35 and 0.33). For this reason, we excluded *sight* from further consideration and construed a metric *othersenses* as a sum of *hearing*, *touch*, *smell* and *taste*, weighted by their respective means. It was scaled so that its mean became the same as that for *sight* (mean=0.79). A similar reasoning determined that out of the 7 measures in *activities*, only *painting* (r=-0.39) and *music-making* (r=0.74) should be considered as they were the strongest explanatory variables, and in two different directions, i.e. one towards *adm,* the other towards *mus*. Being measurements of time, the logarithms were eventually used. The mean cartoon specialist *knowledge* had not been found to be significantly different between groups (Welch two-sample t(25.1)=1.45, p=0.15), but for this purpose it was still considered meaningful. By contrast, certain other data such as *gender* and

CMIO were put aside, despite being more strongly correlated to *group* than *knowledge*, because they did not seem relevant as part of a metric for what is essentially a kind of aural-visual self-image, using a term from acoustic ecology (Truax 1984). More importantly perhaps, *knowledge* correlated negatively with *sight* (r=-0.14) and positively with *othersenses* (r=0.10). While its effects are small, it would contribute to making the division between groups clearer. Finally, the values for Pearson's r between *group* and *othersenses, painting, music* and *knowledge*, respectively, were used as weights to create a summed measure for the difference between the *adm* and *mus* groups:

```
0.52*othersenses+0.74*music-0.25*knowledge-
0.40*painting
```

Figure 6 shows the degree to which individual subjects adhere to the two groups. Scaled to [-1..1] and labeled *avo*, it was taken as an estimate of aural-visual orientation. Figure 6 shows the model's fit with the measures used in its calculation. The different size of the weightings is unsatisfactory, and might indicate a bias, either from the wrong measures being included in the linear combination, or from the limited representativeness of the groups, or from the test questions being inefficient as indicators of aural-visual orientation. Further research based on a stronger theoretical framework will attempt to resolve this question.

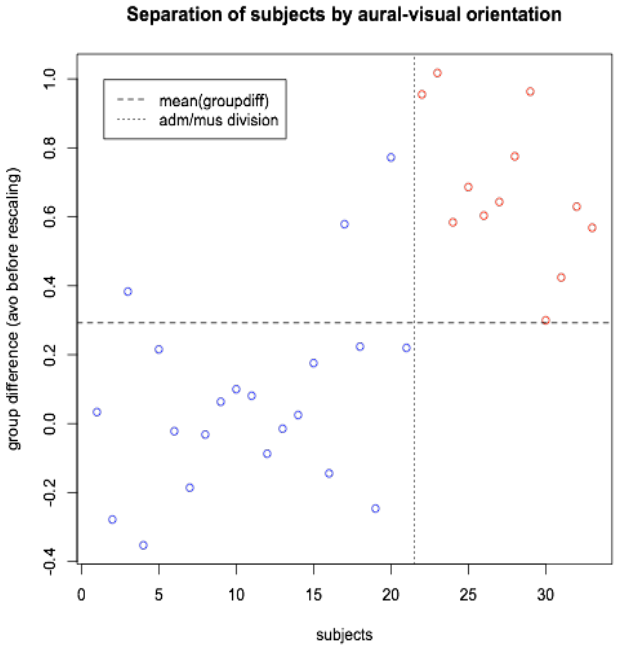

Figure 6. Group adherence "smeared out" by the linear *avo* measure. Subjects are colour coded according to *group*.

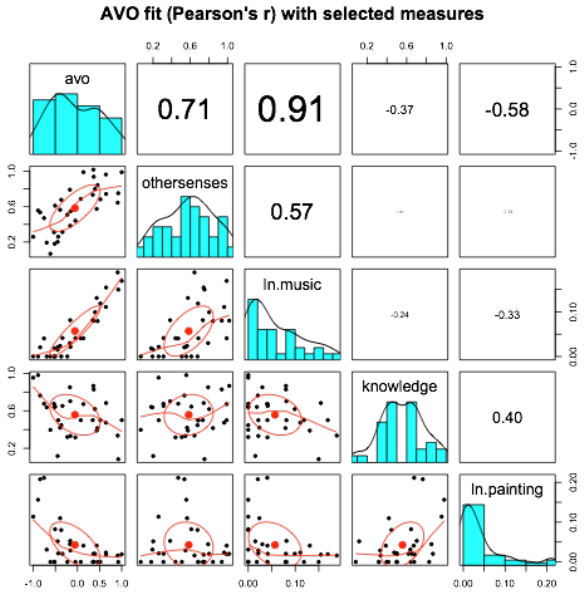

Figure 7. Relative contributions of the selected subject data measures to *avo*.

4.5 *Avo* as a predictor

We looked again at the responses to *dominant* medium (sound or video) in the *both* condition. and found that *avo* could be used as a predictor ($t(337)=-3.24$, $p<0.0014$), i.e. an aurally oriented subject showed a tendency to find "sound" dominant, while someone more visually oriented would rather indicate "video". At the same time, there was a positive prediction from *avo* to *degree,* indicating that aurally oriented subjects found medium dominance (more often *sound*) generally stronger than what the visually oriented did (who indicated a more balanced evaluation). The analysis showed no interaction effect (i.e. *dominant*degree*) correlated with *avo.*

We then considered the unimodal conditions. In all conditions, the *mus* group spent less time on question panels than the *adm* group did, so we looked at the difference between the time each subject spent on responding to *soundelements* and *videoelements* question panels, respectively. One may speculate that the more information a subject takes in and digests from a stimulus, the longer time s/he will take to produce a response. For example, a person who is visually oriented would spend somewhat more time on questions related to visual perception, than on questions about sonic perception. However, it must be underlined that as a causal effect is not testable post-hoc, our analysis could only provide evidence towards the assumption of a causal link being mistaken or not, but not towards it being correct. Taking the difference scores between

the time spent on *soundelements* (in the *sound* condition) and the time spent on *videoelements* (in the *video* condition) produced a measure *svdiff*. We investigated whether it could be predicted by *avo*, and found a significant correlation ($F(1,988)=4.93$, $p<0.027$). It should be mentioned that no significant t test *group* difference between means was found for *svdiff*. This indicates that *avo* is a more discerning measure than *group* on its own. The result may be interpreted as saying that the amount of time people spent on solving the specifically aural and visual tasks in the study had a significant correlation with an independently construed measure for aural-visual orientation.

5. CONCLUSIONS

The article has described the development of audiovisual stimuli characterised by certain emotions. The Cartoon Emotion Experiment results broadly showed a consensus of emotion perception between pre-test judges and CEx test subjects, and consistency across modus conditions by the two groups. The perceived sense of realism was higher in stimuli of negative valence (anger, sadness, fear) than in those of positive valence (joy, love/tenderness). Our data indicated that sound may contribute to making cartoons be perceived as more abstract, which goes contrary to assumptions made in (Cohen 2001).

We also developed a measure for aural-visual orientation, *avo*, as a linear combination of data subjects reported prior to being exposed to the stimuli. It successfully predicted a subject's evaluation of whether sound or video was the dominant medium, and the difference in time s/he spent on solving aural or visual cognitive tasks. While the results are encouraging, they should not be over-interpreted, and are not sufficient to support the hypotheses as stated. Further research, with a more varied subject sample in terms of ages and backgrounds, will be needed to investigate whether aural-visual orientation contributes to explaining emotion responses of various kinds. In particular, an improved measure, even if based entirely on psychometrical data, will have to be corroborated by psychophysiological and other measures.

6. ACKNOWLEDGEMENTS

The CEx study and the author's participation at ICMPC11 have been made possible through Tier 1 Grant M52090026, Academic Research Fund, Singapore.

7. ADDITIONAL FILES

The data from the Cartoon Emotion Experiment are available in a file named `cex.csv`. It is a 59 x 2970 matrix, and includes the calculated *avo* estimates. The stimuli, 30 Mickey Mouse cartoon clips of 10 seconds duration, for practical purposes compressed and reduced in size, are also shared.

8. REFERENCES

Barrier, Mike (1971/2002). "An interview with Carl Stalling". Originally in Funnyworld, Vol 13, 1971. Reprinted in Goldmark, D & Taylor, Y (eds). *The Cartoon Music Book*. Cappella Books, Chicago.

Biancorosso, Giorgio (2009). "Representing Selective Attention in Cinema: the Audio-Dissolve (and its Operatic Roots)". ADM Lecture Series, NTU, Singapore.

Bolivar, Valerie J., Cohen, Annabel J. & Fentress, John (1994). "Semantic and formal congruency in music and motion pictures: effects on the interpretation of visual action". *Psychomusicology*, spring/fall 1994.

Cohen, Annabel J. (2001). "Music as a source of emotion in film." Chapter 11 in Juslin, P & Sloboda, J (eds.) *Music and Emotion: Theory and Research*. Oxford University Press.

Corbett, John (2002). "A Very Visual Kind of Music". In Goldmark, D & Taylor, Y (eds). *The Cartoon Music Book*. Cappella Books, Chicago.

Juslin, Patrik. & Laukka, Petri (2003). "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" *Psychological Bulletin* 2003, vol. 129, no. 5, 770-814.

Lipscomb, Scott & Tolchinsky, David (2005). "The role of Music Communication in Cinema". In Miell D, MacDonald, R & Hargreaves, D (eds). *Music Communication.* Oxford University Press.

McGurk, Harry & MacDonald, John (1976). "Hearing lips and seeing voices." *Nature* 264: 746-8.

Mickey Mouse in black and white: the classic collection (2002). Buena Vista Home Entertainment. [DVD]

Thomas, Frank & Johnston, Ollie (1981). *The Illusion of Life: Disney Animation*. Disney Editions.

Truax, Barry (1984). Acoustic Communication. Ablex Publishing Corporation. Norwood, New Jersey.