

The utility of arc length for continuous response measurement of audience Responses to Humour

Jennifer Juckel^{*} | Steven Bellman^{} | Duane Varan^{***}**

Abstract

Continuous response measures of viewers' reactions to media stimuli have advantages over retrospective self-report measures (Poels & Dewitte, 2006). However, it is difficult to capture continuous measures using single numbers, such as the mean, velocity (the slope), peak, and area under the curve (Kahneman, 2000). Dynamic models are required to illuminate dynamic interactions between audio-visual presentations and audience responses so as not to lose the rich data that is otherwise lost in static measures aggregated over time (Wang, Lang, and Busemeyer, 2011). This is especially the case for humour, which results from a trajectory over time. First there is the set-up, then there is the pay-off. A mean that averages over both these phases is meaningless. In this article, we propose a new measure for characterizing continuous responses: arc length, which is a measure of the length of the curve, or the length of the journey that a measure undergoes during a certain time period.

Keywords: humour, psychophysiology, media, psychology of humour.

* Jennifer Juckel, General Practice Training QLD jjuckel@gptq.qld.edu.au

** Steven Bellman, Ehrenberg Bass Institute, University of South Australia
Steven.Bellman@marketingscience.info

*** Duane Varan, CEO MediaScience varan@mediasciencecelabs.com

Dimensions of humour enjoyment: intensity and valence

Throughout history a variety of physiological measures have been used to mark emotional arousal and attention in media-users. Psychophysiological measures are appealing to researchers as they offset many of the weaknesses associated with collecting subjective data alone, such as reliance on language and memory, temporal imprecision, and interruption of the experimental process (Ravaja et al., 2005; Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006). However, despite these advantages, psychophysiological measures can lead to interpretative ambiguity. As a result, studies typically use a variety of measures together.

A most effective way to measure participant response to media is to measure both arousal and valence (to identify the direction of arousal). According to Wang et al. (2011), a dimensional approach in emotion research is most effectual to capture audience response, whereby valence is represented by the activation of appetitive or aversive systems. Heart rate is a commonly used measure of arousal, however, as the heart is dually innervated it is prone to interpretative ambiguity. SC offers a less ambiguous, simply attained measure of arousal. To measure valence, facial EMG yields reliable results, however at the time of data collection a practical method of measurement was not available. As a result, CRM (dial) was chosen as an effective moment-to-moment measure of valence, which has been shown to be useful in studies involving humour and structural techniques.

Previous ways of characterising continuous responses

Tracing emotional processing over the course of an audio-visual presentation is important when deciphering how responses change from moment-to-moment (Abeele & MacLachlan, 1994). This is especially important for researchers looking to identify differences between responses to individual scenes. Continuous response measurement (CRM) is typically recorded via a handheld dial or slider set to distinguish between increments on a scale, usually some form of semantic differential (Potter & Bolls, 2012). CRM is also low in reactivity, nonverbal, immediate, spontaneous, reliable and valid (Abeele & MacLachlan, 1994).

CRM records the dynamic variation in participant response during a viewing session via introspective analysis, whereas psychophysiological measures do not require such introspection (Potter & Bolls, 2012). As a result, CRM is beneficial for identifying specific subjective responses to particular parts of a presentation, in contrast to facial EMG, where responses are inferred from facial muscle activity. CRM also offers more detailed information than post-exposure retrospective methods of enquiry, where it is unclear which responses are being reported; for example, the peaks

during the presentation, the globally dominant emotion, or the emotion experienced at the end (Fenwick & Rice, 1991; Abeele & Maclachlan, 1994).

Overwhelmingly in the past, mean measures have been used to capture audience response. Woltman Elpers, Mukherjee, and Hoyer (2004) analysed the timing of when peaks in CRM occurred for different groups rating humour and surprise. Their analysis was based on previous work by Baumgartner et al. (1997) who investigated the various properties of CRM traces, including the slope (velocity), change in slope (acceleration), and peak. More recently, Teixeira, Wedel, and Pieters (2012) found that the velocity of changes in facial smiling predict when an ad will be skipped. More generally, Wang et al. (2011) criticized previous analysis of means as confounding the potential causes of moment-to-moment responses, which include not only the currently displaying stimulus on the screen but feedback from previous exposure. All in all, despite their popularity, mean measures are likely to drastically misrepresent the set-up pay-off journey of humour scenes.

Arc length

By showing the journey undertaken by a CRM trace, arc length illustrates some of these feedback effects. Arc length captures variance, the up and down, in a single measure, whereas otherwise three measures, velocity and change in velocity, and change in change in velocity, would be needed to describe the same movement. Because using the mean alone as a measure of continuous responses misses much of the information contained in such rich data, the use of another measure that captures variance within scenes was required. Arc length of each scene was used to capture the journey of peaks and troughs made by each participant's responses. Figure 1 illustrates the difference between long and short arc length in arousal (SC) levels. The first diagram depicts a short arc length, with very little up and down movement, while the second depicts a wavelength with more peaks and troughs, indicating a longer distance (more variance). Two participants could have the same average SC for a scene, but very different arousal journey lengths.

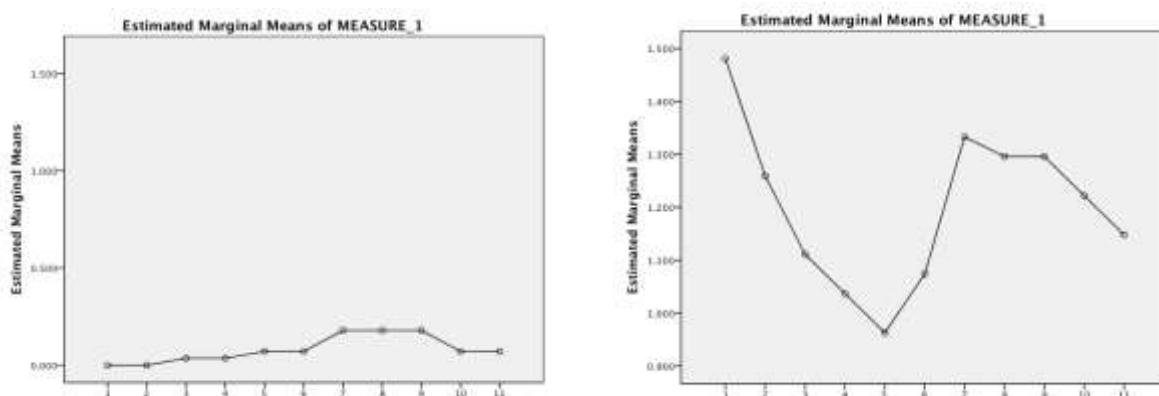


Figure 1 Example of short and long arousal journey (arc length)

Analysing response to humour

In order to work with arc length, a time unit is required that sufficiently captures audience response over time while taking into account the temporal resolution of a single change in SC (6 seconds). This task appears straightforward when analysing a 30-second advertisement, but when analysing a 20-minute sitcom it proves more difficult. For this reason, it makes sense to divide each sitcom into scenes prior to analysis. As sitcom scenes are, in general, around 60 seconds, they are ideal for a time unit equal to the length of each scene (1 minute, on average) divided into 10 sections, which are on average 6 seconds in duration.

Causes of differences in arc length

Wang et al. criticise merely descriptive analyses of CRM data. Accordingly, we proposed that arc length would vary systematically depending on what humour techniques were being used in a scene. To identify these humour techniques, a theoretically exhaustive and mutually exclusive typology of humour techniques was sourced from Berger (1993). This typology, which had originally been developed for the analysis of jokes, had been previously adapted for the analysis of humorous commercials by (Buijzen & Valkenburg, 2004). Pretests using different sitcoms to the ones analysed in the main study demonstrated the need to adapt this typology for use with sitcom humour. The final typology of humour techniques is listed in Table 1.

Table 1 New typology for sitcoms – 4 categories/22 techniques

Category/technique	Short Description
Language	
Allusion	Indirect reference
Irony	Saying one thing and meaning something else or exactly the opposite of what you're saying
Pun	Playing with the meaning of words
Repartee	Verbal banter, usually in a witty dialogue
Ridicule	Making a fool of someone, verbally or nonverbally
Wit*	Ingenious humour
Logic	
Absurdity	Nonsense, a situation that goes against all logical rules
Coincidence	A coincidental and unexpected occurrence
Conceptual surprise	Misleading the audience by means of a sudden unexpected change of concept
Outwitting	Outsmarting someone or the establishment
Caught out*	Unexpectedly get caught while wrongdoing or saying something reprehensible
Misunderstanding	Misinterpreting a situation
Identity	
Parody	Imitating a style or a genre of literature or other media
Rigidity	Someone who thinks along straight lines, who is conservative and inflexible
Malicious pleasure†	Taking pleasure in other people's misfortune; victim humour
Condescension*	Displaying arrogance by patronising those considered inferior
Deceitful behaviour*	Being deliberately misleading, concealing or distorting the truth
Self-deprecation*	Expressing something negative about oneself
Action	
Clumsiness†	Lacking dexterity or grace
Peculiar face†	Making a funny face, grimace
Peculiar music†	Funny, unusual music (when not as part of program structure)
Repulsive behaviour	Offensive, aversive, disgusting behaviour

*New technique created for this study

†Buijzen & Valkenburg technique

Method

Stimuli

This study analysed secondary data collected for another purpose (program testing) by an audience research laboratory in the United States. The original study sampled the highest-rating programs on the four primary US networks. This study analysed a sub-set of the data, the four highest-rating comedy programs across the four primary US networks (ABC, CBS, NBC, FOX). Table 2 displays program and network information. Each stimulus presentation was 60-minutes long and contained two episodes of one program interspersed with four ad pods. Ad breaks were standardized across all programs so that all participants were exposed to the same brands. The advertisements in the pods ranged in duration from 15 – 60 seconds.

Participants

The final sample of participants comprised 62 females and 47 males ($N = 109$). Ages ranged between 19 years – 50 years. The imbalance of males to females arises from the gender split contained in the original dataset, which reflected the intended population of the study. Despite this split, only *Modern Family* (ABC) was affected (females = 19, males = 10), with the remaining three shows containing more even gender numbers; *Big Bang Theory* (CBS, females = 15, males = 12), *Family Guy* (FOX, females = 15, males = 13), and *The Office* (NBC, females = 13, males = 12).

Design

Participants were randomly assigned into one of 4 rotations. The experimental design was between-subjects, with participants viewing two episodes of one program, with advertisements between and throughout the episodes. Each viewing session lasted approximately 60 minutes, followed by a post-session computer survey that lasted between 15 and 20 minutes.

Procedure

Participants watched the stimuli in a small-group (theater) setting. On completion of consent forms, participants were escorted to their assigned viewing theatre to take their seat, each equipped with a computer screen and biometric measurement capability. The maximum group size was nine. An assistant attached electrode to the participant to monitor skin conductance and heart rate (except where a medical condition prevented this). Participants were then presented with a short tutorial on how to use the dial measure of real-time enjoyment (i.e., both dial and biometrics were collected from the same person). The experimental presentation then began automatically. Participants evaluated two recent episodes of television sitcoms while response data during the entire session

were recorded. When the viewing session was over, the assistant returned to remove the biometric leads. Participants were asked to complete the survey on the computer screen on their desk.

Measures

Humour techniques. Sitcoms were first segmented into scenes by topic. The beginning and end points of each scene were coded as markers into a timeline used to analyse the biometric data. Each scene was coded for the presence or absence of the 22 humour techniques, by two independent coders (Krippendorf's alpha > .67)

Arousal. Psychophysiological arousal, which is a continuous measure of a participant's preparation for approach or withdrawal (Cacioppo, Tassinary, & Bernston, 2007), was measured by skin conductance (SC) using disposable SC electrodes. The data were recorded with a Biopac MP35 using AcqKnowledge software (V.4.1, Biopac, USA). Prior to analysis, the SC waveform was filtered to remove high frequency artefacts caused by movements. To control for individual differences, second-by-second SC was converted from raw microSiemens into individual Z-scores. As a result, the transformed data indicated each participant's amount of change, in standard deviations, from their own individual mean throughout the viewing session. Six files were excluded due to response-affecting medical histories or medication use (5.5% of total sample, 109), and 24 files contained bad data (22%). This left a total of 83 usable files. The necessary deletions did not adversely affect the gender split in the SC subsample (females = 46, males = 37), nor the age range (19 – 50 years).

Enjoyment. A continuous real-time measure of enjoyment was captured using a dial that indicated positive or negative valence. The dial set at zero indicated neutral liking, and extended to +3 (extreme positive) and -3 (extreme negative), with increments of 1. As with the SC data, second-by-second dial ratings were transformed into individual Z-scores to control for individual differences in responsiveness.

Arc length. Instead of using calculus to measure the true arc length, which would require customized calculation of derivatives, a very close approximation was calculated using Pythagoras's theorem ($z^2 = x^2 + y^2$). The baseline segment (x) was always defined as one unit in length, although the actual length in seconds could vary. The time unit used to calculate arc length was the length of each scene (1 minute, on average) divided into 10 sections, which were on average 6 seconds in duration. Six seconds is around the temporal resolution of a single change in SC, which takes place approximately 2 to 5 seconds after stimulus onset (Wang et al., 2011). The height (y) was the change in response over the duration of the segment. Enjoyment journey (arc length) was calculated in the same manner as arousal (SC) journey, using the dial data.

Survey. After viewing the program, participants completed a survey that included program evaluation questions (e.g., “How would you rate your enjoyment of this episode?”) on a 7-point scale. Other questions pertaining to attitudes and feelings towards characters and the shows themselves were answered on a 10-point scale.

Covariates. Tests were conducted on the four shows for potential covariates on subsequent analyses. Demographics tested were gender, age, occupation, education, and income. Results revealed no significant differences between groups, indicating successful random assignment. However, because participants were randomly allocated to shows, rather than choosing a show they would like to watch, program fandom was used as a covariate. Participants were asked whether they were a fan of their assigned show on the post-exposure survey (“How much of a fan are you of this show?”). The answer was recorded on a 10-point scale (1 = “Not a fan” to 10 = “Super fan”).

Analysis

The final data set contained 169 rows (the scenes from all 8 episodes), which were the object of analysis rather than participants. Contained in each row were humour techniques information, survey data, and biometric data. A multiple regression analysis using stepwise backward elimination was used to investigate whether individual humour techniques significantly affected scene-level arousal mean, enjoyment mean, arousal journey (arc length) and enjoyment journey measures. The shows used in the experimental presentation, and the mean fan ratings provided by participants on the post-exposure survey were used as control variables.

Results

Dependent Variables: Self-reports and biometrics correlation analysis.

Table 2 Correlations among variables and descriptive statistics

	Fan mean	SC mean	Dial mean	SC arc length	Dial arc length
Fan mean	—				
SC mean	.14	—			
Dial mean	.71 **	.04	—		
SC arc length	.06	.35**	.04	—	
Dial arc length	.22 **	.17*	.36 **	.42 **	—
<i>Program means</i>					
<i>Big Bang Theory</i> (<i>n</i> = 31)	7.27 †	.005 (.032) ^a	(.23) ^b	10.14 (.14 ^b)	10.15 (.11) ^c
<i>Family Guy</i> (<i>n</i> = 42)	6.73 †	.001 (.043) ^a	.55 (.24) ^a	10.09 (.06 ^a)	10.12 (.09) ^{b c}
<i>Modern Family</i> (<i>n</i> = 45)	7.34 †	-.001 (.023) ^a	(.21) ^b	10.07 (.08 ^a)	10.08 (.07) ^{a b}
<i>The Office</i> (<i>n</i> = 51)	7.33 †	.003 (.024) ^a	(.28) ^c	10.05 (.06) ^a	10.04 (.06) ^a
TOTAL (<i>N</i> = 169)	7.17 (.26)	.002 (.031)	1.01 (.37)	10.01 (.09)	10.10 (.09)
Pearson correlations					
** <i>p</i> <0.01 level (2-tailed)					
†Fan mean is a constant for each show					
In any column, means with different superscript letters are significantly different (<i>p</i> <.05), according to a Tukey HSD test.					

Table 2 displays correlations between the dependent variables used in the analysis. The continuous self-report measure of program enjoyment, Dial mean, was significantly correlated with program fandom (Fan mean), as would be expected. The four dependent variables, however, display discriminant validity, as they were not highly correlated (largest $r = .42$ between Dial arc length and SC arc length). Therefore it can be assumed that all four dependent variables are needed, as each contributes unique data pertaining to different aspects of the viewing experience.

The highest correlation in Table 2 is between the two arc length measures, suggesting that dial enjoyment ratings tapped arousal as well as valence (i.e., extreme valence in either direction was associated with high arousal). Dial mean was negatively correlated ($r = -.36$) with Dial arc length.

Because a longer enjoyment journey is caused by positive and negative peaks, negative dips would lower the average over the scene and therefore contribute to the negative association between enjoyment journey and Dial mean. However, there was a medium positive correlation between SC arc length and SC mean. This can probably be explained by the different nature of the dial and SC measures. Dial mean has a neutral zero point at the mid-point of its scale (-3 to +3), whereas SC has a natural zero at extremely low arousal, and participants would tend to rebound to their baseline zero level after arousing events. So while Dial arcs are characterised by peaks and valleys, SC arcs would tend to be lengthened mainly by peaks, and therefore longer SC arc length tends to increase the SC mean for a scene.

Table 3 Correlations among variables and ratings and survey data

	Ratings	Enjoyment mean	Intent to view mean
Ratings	—		
Enjoyment mean	.05	—	
Intent to view mean	.58**	.83**	—
Fan mean	.05	1.00**	.83**
SC mean	.02	-.007	.005
Dial mean	-.18*	.40**	.31**
SC arc length	.36**	-.23**	.01
Dial arc length	.41**	-.29**	-.02
<i>Program means</i>			
<i>Big Bang Theory</i> (n = 31)	12.63 ^a	6.24 [†]	6.12 [†]
<i>Family Guy</i> (n = 42)	8.63 ^b	5.27 [†]	5.47 [†]
<i>Modern Family</i> (n = 45)	10.14 ^c	6.51 [†]	6.40 [†]
<i>The Office</i> (n = 51)	6.70 ^d	6.44 [†]	6.08 [†]
TOTAL (N = 169)	9.19	6.13	6.02
Spearman correlations			
**p<0.01 level (2-tailed)	*p<0.05 level (2-tailed)		
† Enjoyment mean and Intent to view mean constant for each show			
In any column, means with different superscript letter are significantly different (p<.05), according to Tukey HSD test.			

Table 3 displays correlations between the dependent variables used in the regression analysis with ratings and retrospective survey data. Fan mean correlated highly with the survey measures Enjoyment and Intent to view, as would be expected. On the other hand, there was a low correlation between Ratings and Enjoyment mean, highlighting the fact that participants were assigned to their program. This is also reflected in the ANOVA results.

SC arc length (Table 2) aligned with Ratings results (Table 3) in terms of the highest and lowest scoring programs, which were *Big Bang Theory* and *The Office*, respectively, suggesting longer enjoyment journeys are related to increased popularity. *Big Bang Theory* scored highest in SC mean, while also scoring highest in Ratings results (Table 3). Program means for Dial arc length (Table 2) also found concordance with Ratings results (Table 3) in terms of the highest and lowest scoring programs, which were *Big Bang Theory* and *The Office*, respectively, suggesting longer enjoyment journeys are related to increased popularity. These results suggest that a key measure for program testing, especially with forced viewing, is Dial arc length, which in this study was highly correlated with actual ratings data in the field, and is much easier to measure than SC.

Regression analysis. Table 4 displays the significant effects of humour techniques on the four dependent variables. In these analyses, *Modern Family* was used as the default program (the constant). The reason for this being that *Modern Family* was considered the most ‘typical’ show out of those used in the study, in that it does not concentrate on any category of humour techniques, and it generally has a more widespread appeal, as it is not catering to a niche target audience. The remaining shows (as dummy variables) and Fan mean were used as control variables. The following sections report the main results for each dependent variable in turn, beginning with SC mean.

Table 4 Regression results for effect of humour techniques on arousal and enjoyment

Predictor	b	β	SE	t	p
<i>1. SC (arousal) mean</i>					
Constant ^a	-.003		.003	-1.081	.281
HT3.6 Outwitting	.017	.164	.008	2.232	.027
HT4.0 Parody	-.015	-.163	.007	-2.215	.028
HT5.2 Self-deprecation	.055	.197	.020	2.707	.008
<i>2. Dial (enjoyment) mean</i>					
Constant ^a	-4.710		.597	-7.892	.000
HT3.0 Surprise	.170	.157	.058	2.924	.004
HT3.6 Outwitting	.191	.147	.062	3.100	.002
Big Bang Theory ^b	.179	.186	.051	3.525	.001
The Office ^b	.255	.314	.046	5.606	.000
Fan mean ^c	.784	.535	.084	9.306	.000
<i>3. SC (arousal) arc length</i>					
Constant ^a	10.419		.166	62.756	.000
HT3.0 Surprise	.045	.176	.017	2.632	.009
HT3.7 Caught out	.055	.172	.021	2.647	.009
HT5.1 Rigidity	-.058	-.173	.028	-2.071	.040
HT5.2 Self-deprecation	.195	.238	.053	3.648	.000
HT6.2 Pun	.088	.224	.025	3.471	.001
Big Bang Theory ^b	.072	.313	.018	3.963	.000
Fan mean ^c	-.052	-.148	.023	-2.228	.027
<i>4. Dial (enjoyment) arc length</i>					
Constant ^a	10.645		.180	59.190	.000
HT5.2 Self-deprecation	.226	.269	.059	3.834	.000
Big Bang Theory ^b	.060	.255	.016	3.638	.000
Fan mean ^c	-.082	-.228	.025	-3.270	.001
β = Standardised coefficient					
^a Modern Family used as constant (coded 1/0)					
^b Other shows used as control variables (coded 1/0)					
^c Fan mean used as control variable					
$R^2 = .172$ (EDA mean), .678 (Dial mean), .378 (SC arc length), .360 (Dial arc length).					
Largest VIF = 1.741 - HT5.1 (SC arc length)					
$DW = 2.376$ (SC mean), 1.454 (Dial mean), 2.214 (SC arc length), 1.752 (Dial arc length).					

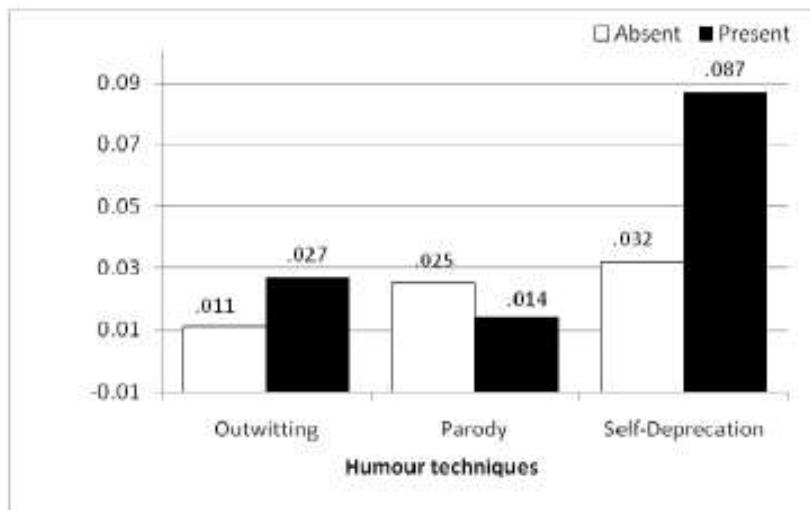


Figure 2 Mean arousal responses to the absence vs. presence of significant humour techniques.

SC (arousal) mean.

Identity. Table 4 shows that three humour techniques were significantly related to SC mean. As can be seen in Figure 2, the Identity category was represented by 2 techniques, Parody and, most significantly, Self-deprecation. Interestingly, this effect of Parody was negative, with the presence of the technique resulting in reduced arousal. Parody was featured most notably in Family Guy, the program to receive the lowest reported enjoyment in this study by way of Dial mean and survey Enjoyment mean. Another category to make an impact on mean SC levels was the Logic category with one humour technique, Outwitting.

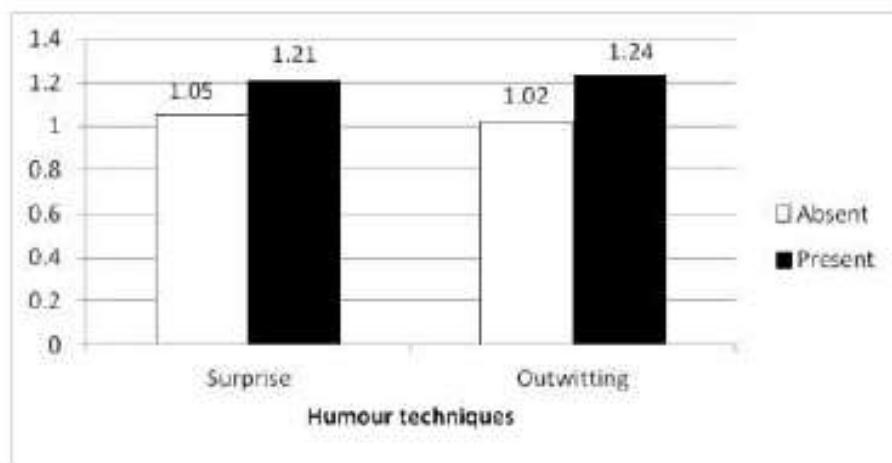


Figure 3 Significant mean dial responses (range -3 to +3) to humour techniques.

Dial (enjoyment) mean. Table 3 shows that two humour techniques were significantly and positively related to Dial mean. As can be seen in Figure 3, the Logic category was the only category to affect Dial mean levels, represented by two techniques. These were Surprise and Outwitting.

SC arc length (arousal journey). Table 3 shows that five humour techniques significantly affected SC arc length, which as shown above appears to be highly predictive of ratings. Of the five techniques that affected arousal journey, Self-deprecation was the most significant ($p = .000$), with presence versus absence positively increasing arc length (peaks and valleys), as can be seen in Figure 4. Also scoring significantly from this category was Rigidity, however the presence of this technique had an inverse effect, reducing arc length. This effect was not highly significant ($p = .04$) and is difficult to perceive in Figure 4. Techniques from the Logic category also significantly affected arousal journey, with Surprise and Caught out being highly significant techniques ($p = .009$). As shown in Figure 4, the presence of either of these variables positively increased SC arc length.

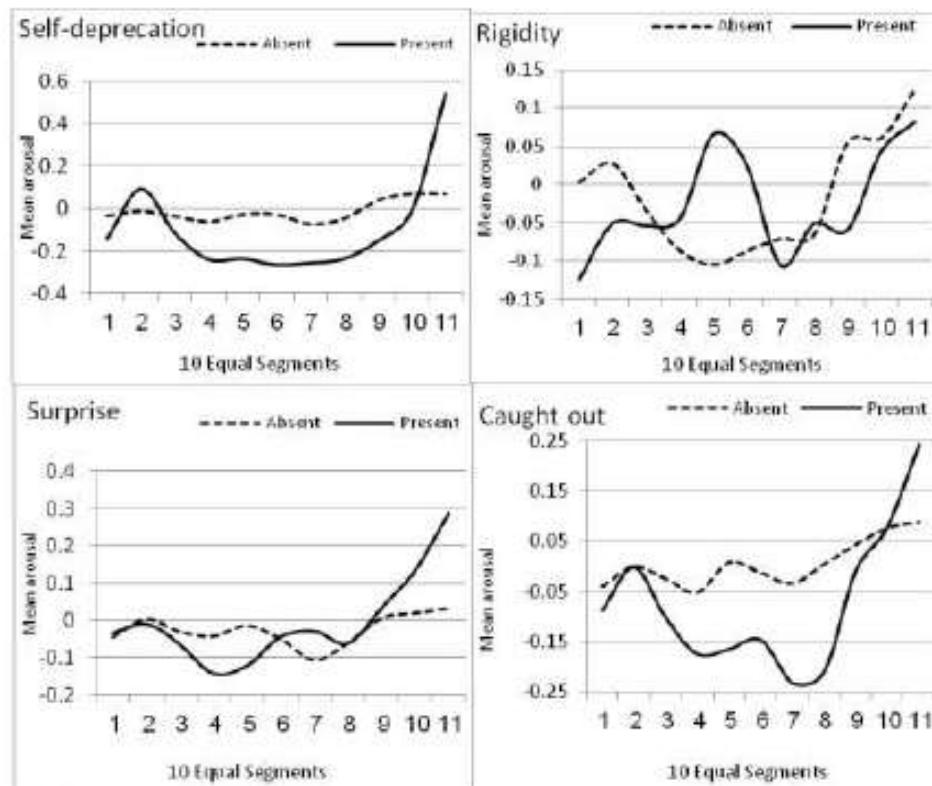


Figure 4 Arousal journey (arc length) for significant humour techniques

As can be seen in Figure 5, the Pun technique from the Language category also had a strong significant positive effect on arousal journey, and was the only technique from this category to affect arousal.

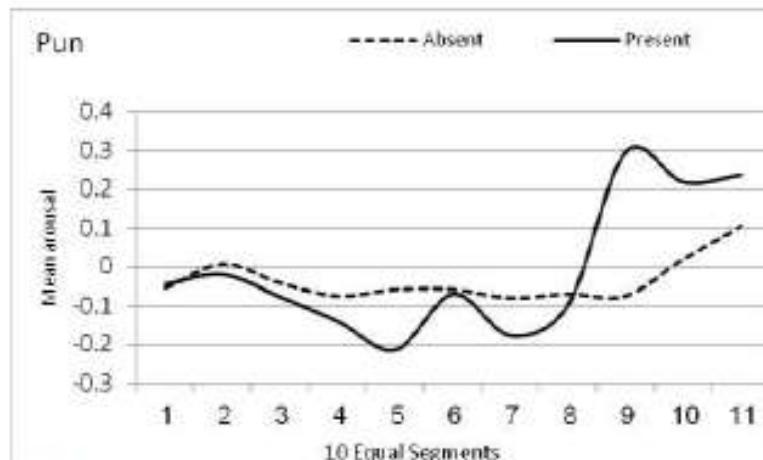


Figure 5 Arousal journey (arc length) for humour technique Pun

Dial arc length (enjoyment journey). Table 3 shows that one humour technique is significantly related to Dial arc length (enjoyment journey). The Identity category's Self-deprecation humour technique had a positive effect on enjoyment journey, as can be seen in Figure 6.

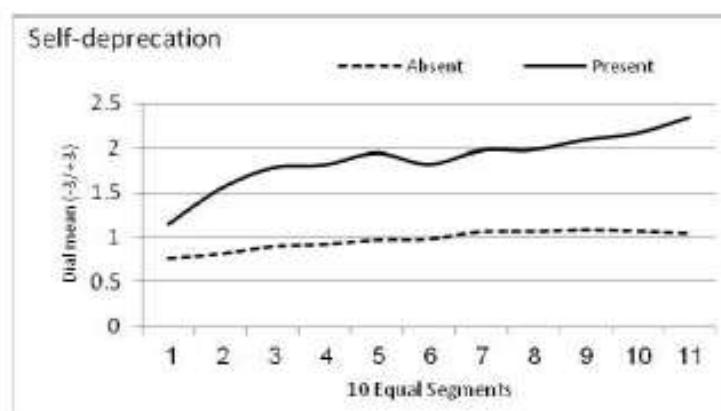


Figure 6 Enjoyment journey (arc length) for humour technique Self-deprecation

Discussion

Superiority of arc length measures

Two types of dependent variable were used: (1) the mean value of a variable across the duration of a scene, and (2) the journey or arc length of the curve traced by the variable over duration of a scene. Emotional journey was expected to provide more information than a traditional means analysis, because arc length includes information about the variance in a variable over a scene, that is, the journey it makes above and below its mean value.

The lack of relationship between means and journey length shown in the results justifies the use of both types of measure. However, arousal (SC) journey was affected by more of a variety of techniques than the other dependent variables, with five humour techniques from three categories being represented.

Analysis also revealed that dial and SC arc length (journey) measures were more predictive of program performance than traditional mean measures, with stronger correlations between ratings and these measures. Perhaps these measures tap the reactions that are an essential part of successful sitcoms. It is interesting that dial journey, like SC journey, was predictive of ratings, even though, unlike SC, dial journey requires conscious ratings by the viewer. But dial journey was less sensitive than SC journey to the effects of different humour techniques.

Practical Application

Analysis revealed that SC arc length measures were affected by a wider variety of techniques than dial arc length or either of the means measures. Furthermore, and most importantly, arc length measures were more indicative of ratings success than the means measures. This suggests it would be worthwhile including these measures in any efforts to predict sitcom success (e.g., pilot testing). In addition, the fact arc length measures were affected by a wider variety of techniques indicates that these measures are more sensitive than the traditional means measures, making them a more useful inclusion when only a few measures of enjoyment are practical.

Future Research

Future research should use the humour techniques typology and biometrics measures developed in this study on a wider variety and higher number of sitcoms — both successful and unsuccessful. These results are correlational only, not causal, so research that manipulates the factors identified in the regression analysis, while comparing arc length measures with ratings, would definitively identify the building blocks of a successful sitcom (Armstrong, 2012).

Live performance would be an interesting arena for experimentally investigating the effects of humour techniques on arc length, as live performance is much easier to manipulate than video production (Russell, 2002).

Conclusion

Arc length is a new way of characterizing CRMs that is very promising. Our results show it yields superior insights compared to simple means analysis. Future research is needed but already these results suggest that this new measure has immediate practical utility.

Sitcom success is difficult to predict, posing a problem for all involved with the sitcom industry. To this end, two new experimental measures (of arc length) were found to be more indicative of actual ratings success than traditional means-based measures. These findings are, however, best considered a strong starting point for further research, due to the limitations of the current study. A wider variety of sitcoms and comedy genres should be tested, using the new typology in conjunction with the new emotional journey (arc length) measures, with a view to developing genre-specific prediction kits for use by writers, producers, and television networks.

References

- Abeele, P. V., & Maclachlan, D. L. (1994). Process tracing of emotional responses to TV ads: Revisiting the warmth monitor. *Journal of Consumer Research*, 20(4), 586-600.
- Baumgartner, H., Sujan, M., & Padgett, D. (1997). Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgements. *Journal of Marketing Research*, 34, 219-232.
- Buijzen, M., & Valkenburg, P. M. (2004). Developing a typology of humor in audiovisual media. *Media Psychology*, 6, 147-167.
- Cacioppo, J., Tassinary, L. G., & Bernston, G. G. (2007). *The handbook of psychophysiology*. New York: Cambridge University Press.
- Fenwick, I., & Rice, M. D. (1991). Reliability of continuous measurement copy-testing methods. *Journal of Advertising Research*, 31(1), 23-29.
- Kahneman, D. (2000). Evaluation by moments: Past and future. In D. Kahneman & A. Tversky (Eds.), *Choices, Values and Frames*. New York: Cambridge University Press.
- Poels, K., & Dewitte, S. (2006). How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research*, 46(1), 18-37. doi: 10.2501/S0021849906060041
- Potter, R. F., & Bolls, P. D. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. New York: Routledge.
- Ravaja, N., Saari, T., Laarni, J., Kallinen, K., Salminen, M., Holopainen, J., & Jarvinen, A. (2005). *The psychophysiology of video gaming: Phasic emotional responses to game events*. Paper presented at the DiGRA 2005 Conference: Changing Views: Worlds in Play.
- Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., & Kivikangas, M. (2006). Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence-Teleoperators and Virtual Environments*, 15(4), 381-392.

- Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research*, 49(2), 144-159. doi: <http://dx.doi.org/10.1509/jmr.10.0207>
- Wang, Z., Lang, A., & Busemeyer, J. R. (2011). Motivational processing and choice behaviour during television viewing: An integrative dynamic approach. *Journal of Communication*, 61, 71-93. doi: 10.1111/j.1460-2466.2010.01527.x
- Woltman Elpers, J., Mukherjee, A., & Hoyer, W. D. (2004). Humor in television advertising: A moment-to-moment analysis. *Journal of Consumer Research*, 31(3), 592-598.