AUDIOVISUAL PERCEPTION OF AROUSAL, VALENCE, AND EFFORT IN CONTEMPORARY CELLO PERFORMANCE

Hanna Järveläinen
ICST, Zurich University of the Arts
hanna.jarvelainen@zhdk.ch

ABSTRACT

Perceived arousal, valence, and effort were measured continuously from auditory, visual, and audiovisual cues using a recorded performance of a contemporary cello piece. Effort (perceived exertion of the performer) was added for two motivations: to investigate its potential as a measure and its association with arousal in audiovisual perception. Fifty-two subjects participated in the experiment. Results were analyzed using Activity Analysis and functional data analysis. Arousal and effort were perceived with significant coordination between participants from auditory, visual, as well as audiovisual cues. Significant differences were detected between auditory and visual channels but not between arousal and effort. Valence, in contrast, showed no significant coordination between participants. Relative importance of the visual channel is discussed.

1. INTRODUCTION

Describing emotional response to music, the circumplex model presents a variety of affects as a combination of two dimensions, arousal and valence (see [1, 2] for a review). In tonal classical music, the main cues for high arousal are fast tempo and high intensity, while major mode, fast tempo, and high pitch contribute to positive valence [3]. Western listeners learn cultural valence cues from classic-romantic and from popular music. These cues are often absent in contemporary music, where valence must be judged from acoustic attributes. Valence models are less successful under such circumstances [4]. Dean and Bailes found a potential association between valence and spectral flatness in electroacoustic music [5]. However, the effect was also overruled by higher-level features.

Dean and Bailes also discovered loudness patterns in electronic compositions resembling those resulting from the player's exertion in classical music [6]. They suggested that effort would be a key element in the FEELA-chain leading to emotional arousal (Force, Energy, Effort, Loudness, Arousal). In the auditory domain, effort is associated with intensity, source complexity, and event density in varying degrees, as long as human musical agency is apparent to the listener [7]. In the visual domain, effort is considered a factor of expressive body movement in musical and dance performance. However, perception of effort changes has been less explored.

In music performance, various visual cues have been identified from gestures communicating specific affects such as anger, fear, grief, and joy [8]. General importance of visual kinematic cues has been confirmed for higher-level features such as performance judgment [9], expertise [10], player identification [11], structural phrasing [12, 13], expressive intentions [14–16], and various emotional cues [17–19]. However, basic affect perception from audiovisual cues has received less attention until recently. Vines and colleagues measured continuous perception of musical tension, reporting great differences between the auditory and visual channels [12]. Vuoskoski et al. studied expressivity [20] and emotional impact [21], concluding that visual cues were of equal or even higher importance than auditory cues. Yet the relative importance of visual cues in real-time perception still requires research.

The present experiment is motivated through the recent evidence of the role of the visual channel. Effort was included in order to explore its suitability as a measure in contemporary repertoire and its differences with arousal in audiovisual perception. The cello was chosen for good visibility of the playing gestures and their importance to loudness control. Based on literature and on pilot experiments, it is expected that both arousal and effort will be judged reliably from auditory and from visual cues. The two measures are expected to be positively associated. Suitability of valence as a measure of perceived affect in contemporary repertoire is discussed.

2. EXPERIMENT

2.1 Design, stimuli, and apparatus

Three factors were varied in the experiment: measurement (arousal, valence, and effort), sensory modality (A=auditory only, V=visual only, and AV=audiovisual), and musical material (segments 1-3). The material consisted of three excerpts from an audio-video recording of the solo cello work Pression by Helmut Lachenmann. This piece was chosen because it lacks melodic and harmonic elements as well as regular beats, yet it contains a rich variety of playing techniques, gestures, and timbres. The video recording was made using a professional camera. The performer was filmed from a distance of ca four meters against a still, dark, and neutral background, offering good contrast to her clothing. The audio track was two-channel at 48 kHz. [1] The duration of the segments varied between 2 min and 2 min 30 sec. The unimodal A and V conditions contained only the audio and the video tracks, respectively, while the bimodal AV condition contained both. Audio and video were always congruent.

Subjects made continuous ratings of perceived (not experienced) affect using a slider while observing the performance one factor combination at a time. A custom-made software played back the excerpts on a 13-inch laptop computer and collected responses at a sampling frequency of 4 Hz. Audio was presented through Direct Sound EX-29 Extreme Isolation headphones at realistic and comfortable level, and video was shown in full-screen mode. Subjects were instructed to rate perceived, not experienced affect, and the meanings of the three measurements were explained to them before each measurement block. Arousal was described varying between tense/relaxed or awake/tired. Effort was defined as the musician's exertion in order to produce the sound. For these measures, the slider setting was mapped to the numeric range [0,1]. Valence was described as varying between positive/negative, pleasant/unpleasant, attractive/unattractive, or happy/sad. The slider was mapped to the range [-0.5, 0.5].

2.2 Subjects and procedure

Fifty-two students participated in the experiment (ages 20-45 years, M = 27; 16 M, 36 F). Roughly two thirds were music majors, the rest had no significant musical experience. Participants' musical background was recorded but not controlled as a grouping factor. The session took ca 30 minutes. After completing the experiment, participants reported their liking and perceived familiarity of the repertoire, both on a scale from one to five.

Each participant was assigned nine of the 27 factor combinations as follows. At first, the six permutations of segment numbers 1-3 were assigned to the modalities A, AV, and V, producing triplets such as (A1, AV2, V3). Then, one of these triplets was assigned to each measurement such that the participants received each modality once within the three segments of a single measurement, and each modality once within the three measurements of a single segment. There were 12 such sets; thus each set was received by four or five participants. An example of the conditions received by a single subject is given in Table 1.

This scheme produced 17 ratings per measurement-modality-segment bin (16-18 due to practicalities). There were no common participants in bins with matching modality and segment (for example, effort-A-1, arousal-A-1, valence-A-1), nor in bins with matching measurement and segment (for example, effort-A-1, effort-V-1, effort-AV-1). However, between non-matching factor combinations (such as effort-A and valence-AV), there was up to 50% overlap between participants. The three measurements were presented as blocks in balanced order.

[1] Recording available at https://tube.switch.ch/videos/db27af24

	A	AV	V
Arousal	Seg. 1	Seg. 2	Seg. 3
Valence	Seg. 2	Seg .3	Seg. 1
Effort	Seg. 3	Seg. 1	Seg. 2

Table 1: One of the 12 sets of factor combinations.

2.3 Data analysis

Coordination in participants' responses was investigated using Activity Analysis, a novel analytical framework based on alignment between continuous responses of different subjects [22]. Well aligned responses are probably driven by the stimulus and not produced randomly. Activity Analysis begins by searching individual responses for active events in terms of enough change in a given time frame. Activity levels are then computed as the proportion of responses that show a similar kind of event within a given time window of synchrony. Sequenced assessment of activity levels over the duration of the measurement produces the activity level time series. In this study, activity levels are computed from rating increases of at least 2.5% within 2-second time windows. This in turn is used for computing the Coordination Score by testing the distribution of activity levels against a parametric model of uncoordinated random activity. The single-number C-Score varies between 0 and 16, with $C > 2$ indicating significant coordination on a $p < 0.01$ level. A Bi-Coordination Score can be computed between two collections of ratings with different response conditions and participants. These analyses were performed in Matlab using the Activity Analysis toolbox [22].

Functional data analysis was performed in R using the fda.usc package [23]. For this analysis, the ratings were converted into functional data objects and then smoothed using nonparametric kernel estimation. The data were used in original as well as differenced form. To investigate differences in functional means between conditions, functional analysis of variance was performed based on randomly chosen one-dimensional projections [24]. Two-way between-subject ANOVAs were computed for each segment with measurement and modality as factors. [2]

3. RESULTS

3.1 Arousal and Effort

Functional means of combined arousal and effort ratings are presented in the top panels of Figures 2, 3, and 4 [3]. Peak ratings are in all segments reached in the auditory condition. In segments one and three, peak auditory ratings top visual ratings by nearly 20 percentage points and audiovisual ratings by ca 10 pp. Crossmodal additive effects do not seem present; audiovisual ratings never exceed the higher unimodal condition.

[2] The segments were treated as separate experiments, as a comparison of different musical materials would not be meaningful.

[3] Averaging over the two measurements is justified by upcoming analysis.

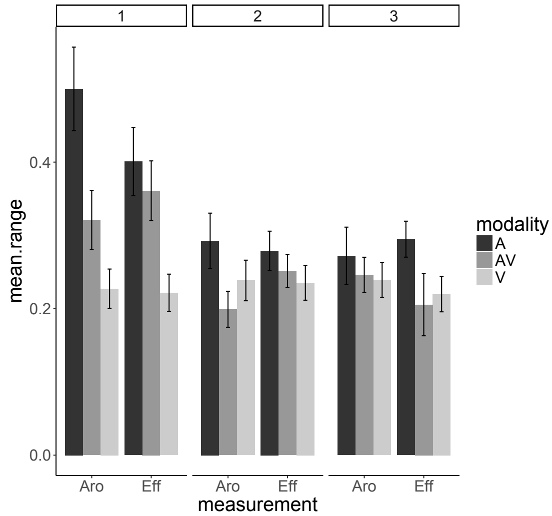

Figure 1: Mean ranges (±se) of the ratings curves in segments one, two, and three, as proportion of the ratings scale.

Participants' individual use of the rating scale was examined as the range between the 25% and 75% quantiles of their ratings. Figure 1 presents the mean ranges across participants. Typically, they used ca 25% of the rating scale, and the ranges were widest in the the auditory condition.

3.1.1 Activity Analysis

Coordination Scores for all factor combinations are given in Table 2. An example of ratings and the respective activity level time series is seen in Figure 5. In nearly all factor combinations, ratings were significantly coordinated ($C > 2$), with an overall mean $C = 3.79$. However, only the auditory modality resulted in significant C-Scores in all factor combinations. The audiovisual modality was uncoordinated in two cases and the visual condition in one.

Bi-Coordination-Scores were computed to compare alignment between measurements and modalities (Table 3). The auditory and visual conditions were uncoordinated with only one exception. In contrast, auditory and audiovisual ratings were coordinated, except for arousal in segment one (mean Bi-C of the coordinated conditions = 3.66). Visual and audiovisual ratings were likewise coordinated, albeit with a lower mean Bi-C = 2.79. Arousal and effort measurements were always coordinated within matching modalities (mean Bi-C = 4.13).

The Activity Analysis results can be summarized as follows. Firstly, participants rated both arousal and effort in a coordinated way from isolated as well as combined auditory and visual cues. Secondly, there was significant bi-coordination between arousal and effort ratings in all modalities. Thirdly, even though there was generally no significant bi-coordination between auditory and visual ratings, both modalities were bi-coordinated with audiovisual ratings, suggesting that audiovisual perception is significantly driven by both auditory and visual cues.

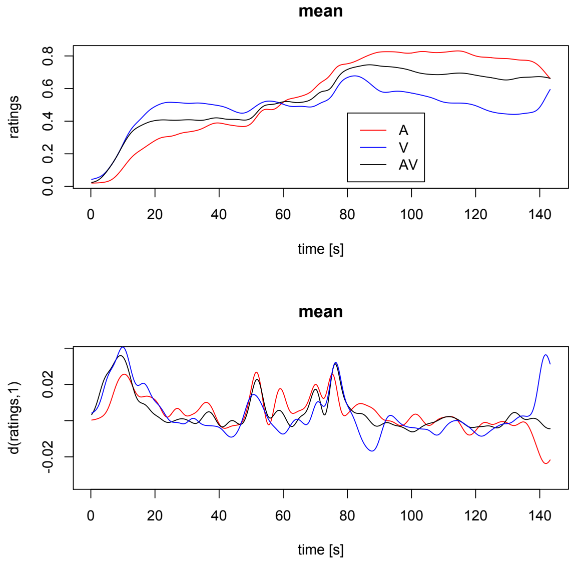

Figure 2: Original (top) and differenced (bottom) mean ratings (arousal and effort combined), segment one.

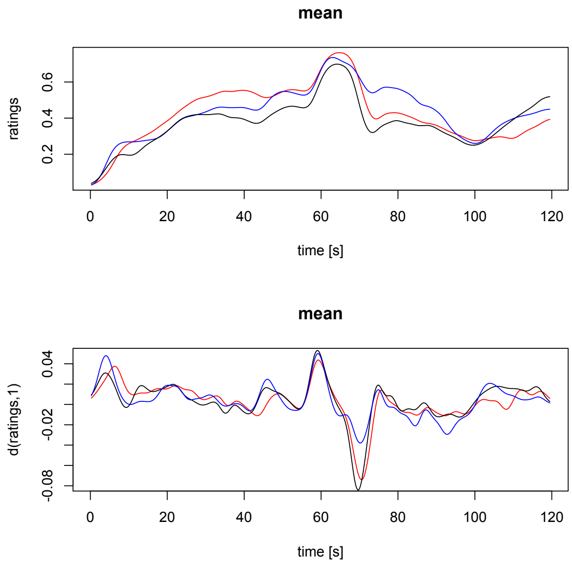

Figure 3: Original and differenced ratings, segment two.

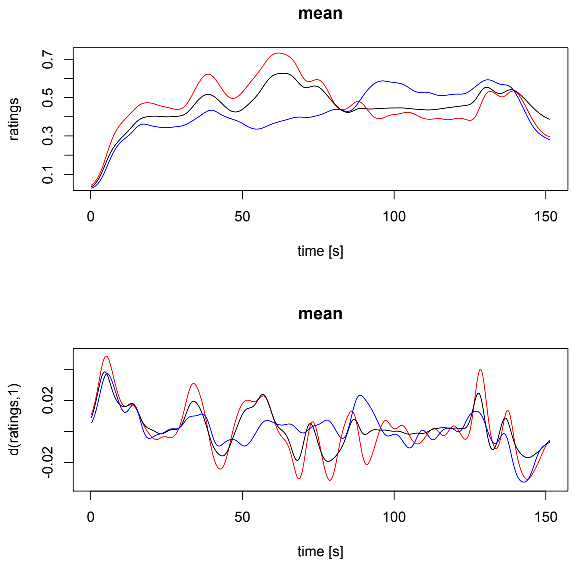

Figure 4: Original and differenced ratings, segment three.

	Arousal C-score	Effort C-score
Seg. 1	A: 3.36 V: 4.13 AV: < 2	A: 4.28 V: 2.80 AV: 4.40
Seg. 2	A: 3.06 V: 4.37 AV: 4.76	A: 4.71 V: <2 AV: 4.76
Seg. 3	A: 2.70 V: 2.29 AV: 9.08	A: 2.42 V: 7.38 AV: < 2

Table 2: Activity Analysis C-Scores.

	Arousal Bi-C-score	Effort Bi-C-score	Aro-Eff Bi-C-score
Seg. 1	A-AV: < 2 V-AV: 2.25 A-V: < 2	A-AV: 2.94 V-AV: 2.25 A-V: < 2	A: 4.13 V: 3.40 AV: 3.68
Seg. 2	A-AV: 2.77 V-AV: 3.47 A-V: < 2	A-AV: 2.74 V-AV: 3.54 A-V: < 2	A: 3.10 V: 3.19 AV: 3.67
Seg. 3	A-AV: 4.67 V-AV: 2.90 A-V: 2.56	A-AV: 5.18 V-AV: 2.35 A-V: < 2	A: 5.65 V: 5.56 AV: 4.84

Table 3: Activity Analysis Bi-C-Scores.

3.1.2 Functional analysis of variance

Functional two-way ANOVA was computed segment-wise with measurement and modality as factors. A significant main effect was observed for modality: the p-value, obtained from 30 random projections, was $p < 0.001$ for all three segments. On the contrary, the measurement effect was not significant in any segment ($p \geq 0.43$), nor was the measurement:modality interaction ($p \geq 0.16$).

Special contrasts were computed, using the Bonferroni method, for all modality pairs. Significance levels for these contrasts are listed in Table 4. The visual ratings always differ significantly from both auditory and audiovisual ratings. In segments one and three, the difference between audiovisual and auditory ratings is non-significant or marginally significant. In segment two however, the audiovisual ratings differ significantly from both auditory and visual ratings. As seen in Figure 3, the visual channel seems to dominate first and the auditory channel thereafter.

3.1.3 Differenced ratings

The functional ANOVA analysis was repeated for differenced ratings, seen in the bottom panels of Figures 2, 3, and 4. The goal was to investigate, whether some of the findings in the first analysis would be due to an off-set rather than a profile difference. A further motivation for analysing differenced data is their reduced dependency on previous values. The results confirm the first analysis: modality had a significant main effect ($p < 0.01$) but measurement did not ($p \geq 0.19$). Nor was there a significant

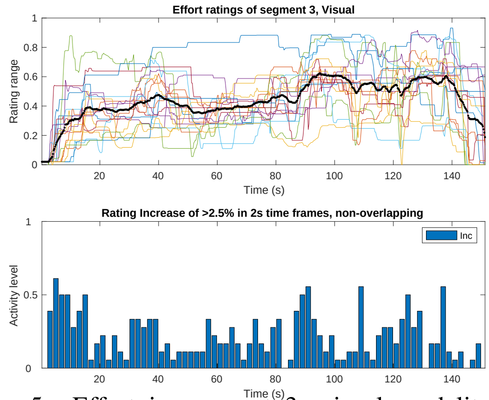

Figure 5: Effort in segment 3, visual modality. Top panel: individual and mean ratings (bold black line), bottom panel: activity levels for rating increases.

Original	AV-A	AV-V	A-V
Segment 1	* ($p = 0.021$)	**	***
Segment 2	***	***	**
Segment 3	-	***	***
Differenced	**AV-A**	**AV-V**	**A-V**
Segment 1	- ($p = 0.059$)	***	***
Segment 2	-	**	**
Segment 3	-	***	***

Table 4: Planned contrasts in functional analysis of variance for original and differenced data. Significance levels: $p < 0.001$***, $p < 0.01$**, $p < 0.05$*, $p > 0.05$-.

interaction ($p \geq 0.48$). Compared to the first analysis, the contrasts between audiovisual and auditory ratings in segments one and two are non-significant. These differences are therefore rather of the off-set type. This suggests that in terms of profile, the audiovisual ratings are closer to the auditory than the visual ratings.

In terms of distance, the (non-differenced) audiovisual ratings are not substantially closer to the auditory ratings or even to the channel whose ratings are higher. Rather, the dominant channel might be the one changing more radically. In this respect, segment two at 60-80 s is particularly interesting. There the ratings drop suddenly in all three conditions; however, audiovisual ratings seem to be captured by the steeper decline in the auditory channel.

3.2 Valence

Valence ratings were mostly uncoordinated. Only one of the nine factor combinations was significantly coordinated (C=3.38 for auditory modality in segment three); for the rest, the C-Score was insignificant (C<2). While the auditory modality reached a mean C-Score not far from significant (C=1.84), the visual and audiovisual modalities were clearly below it (C=0.68 and C=0.83, respectively). An example of the ratings and activity levels in the visual modality is seen in Figure 6.

Figure 6 also shows great individual differences typical of the valence measurement. At the same time as some

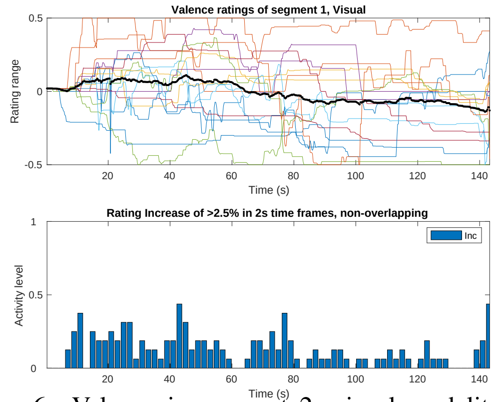

Figure 6: Valence in segment 2, visual modality. Top panel: individual and mean ratings (bold black line), bottom panel: activity levels for rating increases.

participants rated the passage as maximally positive, others gave maximally negative ratings. The resulting mean curve hardly deviates from zero. It is therefore impossible to make further conclusions based on the mean.

The reasons for the differences are of interest, however. As was seen, participants covered varying ranges. There was a positive association between participants' liking scores and mean valence levels taken across all their valence ratings (Pearson's $r = 0.55$, $p < 0.001$). Liking and familiarity were likewise positively associated ($r = 0.51$, $p < 0.001$). However, there was no significant association between mean valence levels and familiarity ($r = 0.21$, $p = 0.14$), nor between mean valence levels and being a musician ($r = 0.09$, $p = 0.51$).

4. DISCUSSION

4.1 Audiovisual perception

Arousal and effort were generally perceived with significant coordination between participants from auditory, visual, and audiovisual cues alike. Although both visual and audiovisual ratings were significantly coordinated in most of their factor combinations, only auditory ratings reached significant coordination in all cases. Participants also utilized a slightly wider range in the auditory ratings. Thus, auditory cues produced somewhat better alignment and higher variability in the rating curves than visual cues.

Significant bi-coordination was generally lacking between the unimodal auditory and visual conditions, and functional analysis of variance detected a significant difference between them in both original and differenced ratings. This outcome is in line with the study by Vines and colleagues, who observed significant differences between auditory and visual perception of musical tension [12]. They concluded that audiovisual ratings were dominated by the auditory modality. Here, such evidence is less conclusive. Even though the audiovisual rating profiles match better with the auditory ratings, the Activity Analysis confirms that both channels significantly drive audiovisual perception.

The theory of optimal sensory integration predicts that bimodal perception is dominated by the modality delivering more reliable information [25–27]. The auditory channel could be labelled generally more reliable, because it was perceived at a wider dynamic scale and thus contains more noticeable changes. One might argue that this difference was due to less realistic visual playback through a small screen. Literature on the effect of screen size on perceived affect is not exhaustive; iPhone-size screens are known to reduce immersion [28, 29]. It is indeed possible that visual perception might have been underestimated here in comparison to a hypothetical live concert situation. Live measurements, as a function of seating and lighting conditions in the hall, would be necessary to estimate the effect. Given that even with the current setup, mean visual ratings exceeded mean auditory ratings ca. 35% of the time, the effect of video presentation should be but small. Also various pilot experiments measuring effort, made using variable bigger screen sizes, delivered results similar to the current experiment. Moreover, one would expect the suppressing effect to be a negative off-set. The current results indicate, however, that the main reason why participants were more tuned to the auditory ratings were profile differences.

Analysis of the profiles in terms of differenced ratings suggests that audiovisual perception may have been captured by change. In this experiment, more change was generally perceived in the auditory channel, thus it transmitted more information than the visual channel. This seems only natural, given that the sound of musical instruments is supposed to respond to even very fine adjustments in playing gestures [30]. Moreover, because this study focused on the perception of basic affects, the material did not contain additional expressive gestures typical of traditional cello performance. If added, these might significantly increase the amount of movement, which could have an effect on perceived arousal along with other emotional cues. The current measurement therefore addresses a baseline of perceived audiovisual arousal, which could be generalized to both contemporary and classical cello repertoire containing mainly sound-producing movements. As a further step, the effect of expressive gestures on perceived arousal would be of interest.

As an immediate next step, visual attributes underlying arousal and effort perception is a research interest following this study. Intensity, and to some extent spectral centroid and spectral flatness, are known auditory cues for arousal [5,31,32]. Visual cues, movement size particularly, were found to convey loudness but not tempo changes in piano performance [21]. The first step is to investigate, how much of the variation in auditory and especially visual ratings is explained by intensity. In the visual domain, this should depend on the way loudness is controlled in an instrument. In bowed string instruments, loudness is increased through bow velocity, requiring larger and faster movements. The mapping between movement size and intensity is obvious also in the piano, but less so in wind instruments, the organ, and the harpsichord. Would the visual channel transmit even fewer bits of information in these cases, and lose importance? Or would the judgment require expertise on the instrument?

4.2 Arousal vs effort

Arousal and effort were bi-coordinated in all factor combinations, and no main effect of measurement was detected by functional analysis of variance. It is probable that participants judged both from the same cues, the main difference being the viewpoint. One of the goals of this study was to investigate, whether effort and arousal are perceived differently through auditory and visual channels. While exertion can be relatively well estimated from visual cues [33, 34], it was hypothesized that judging it from auditory cues might require more inference and be more difficult. Such differences were not observed. Non-musicians perceived on average nine percentage points more effort than musicians in the visual condition. However, the nature and significance of this potential effect cannot be estimated from the current data.

According to the FEELA hypothesis, the player's exertion is transmitted through intensity changes to perceived arousal [5, 7]. This implies that effort cues must also be arousal cues. Whether the contrary is true must depend on the circumstances. As long as the intensity changes are caused by the performer's exertion, effort and arousal perception should match well. This seems evident in the present case. Further experiments could be designed using music where intensity changes do not originate from effort changes, such as looming motion, or music with no intensity changes.

4.3 Valence

The valence measurement lacked significant coordination. This was not surprising, given the absence of culturally learned cues. There was a significant positive association between participants' individual mean valence level and their self-reported liking of the repertoire. Furthermore, valence ratings predicted the liking scores in two segments. Thus, subjects' attitude may have influenced their perception of one of the two basic affect dimensions. The influence of mood on perception has previously been observed using visual emotional stimuli [35]. Hence it is concluded that valence was not a reliable measure of emotional response to the repertoire in this study. However, detailed analysis of the valence ratings may reveal moments of high activity, even if the general coordination scores were low. Such moments may be of interest considering covariance with the other two measures, and will be analyzed at a later stage.

5. CONCLUSIONS

Perception of basic affect dimensions from auditory and visual cues was measured in cello performance. While the auditory and visual channels were perceived differently, ratings were generally significantly coordinated within each modality. Arousal and effort were perceived similarly, supporting the notion that in acoustic music performance, both are associated with loudness changes.

Auditory and visual ratings of arousal and effort were different but both influenced audiovisual perception. It is hypothesized that subjects' attention was caught by the more dramatically changing modality. In the present case, more change was perceived in the auditory channel. This might be due to contents of the visual channel; however, an additional effect of transmission medium (screen size) cannot be eliminated. However, if the visual component is present in either live or recorded performance, both auditory and visual channels should ideally communicate a roughly equal amount of information in order not to suppress audiovisual perception. The present study concerns basic affect perception; in literature, visual cues have been shown to communicate higher level features, such as expressiveness or expertise, perhaps even more reliably than auditory cues.

Valence perception was uncoordinated between participants. Individual mean valence ratings were associated with participants' liking of contemporary music.

Acknowledgments

Jan Schacher and Ellen Fallowfield are gratefully acknowledged for producing the musical material used in this experiment.

6. REFERENCES

[1] M. Zentner and T. Eerola, "Self-report measures and models," in *Handbook of Music and Emotion: Theory, Research, Applications*, P. Juslin and J. Sloboda, Eds. Oxford University Press, 2010.

[2] P. Juslin and R. Timmers, "Expression and commununication of emotion in music performance," in *Handbook of Music and Emotion: Theory, Research, Applications*, P. Juslin and J. Sloboda, Eds. Oxford, UK: Oxford University Press, 2010.

[3] I. Peretz, L. Gagnon, and B. Bouchard, "Music and emotion: perceptual determinants, immediacy, and isolation after brain damage," *Cognition*, vol. 68, no. 2, pp. 111–141, 1998.

[4] K. N. Olsen, R. T. Dean, and C. J. Stevens, "A continuous measure of musical engagement contributes to prediction of perceived arousal and valence." *Psychomusicology Music. Mind, Brain*, vol. 24, no. 2, pp. 147–156, 2014.

[5] R. T. Dean and F. Bailes, "Time series analysis as a method to examine acoustical influences on real-time perception of music," *Empir. Musicol. Rev.*, no. 4, pp. 152–175.

[6] R. Dean and F. Bailes, "A Rise-Fall Temporal Asymmetry of Intensity on Composed and Improvised Elecroacoustoc Music," *Organ. Sound*, vol. 15, no. 2, pp. 147–158, 2010.

[7] K. N. Olsen and R. T. Dean, "Does Perceived Exertion Influence Perceived Affect in Response to Music? Investigating the FEELA Hypothesis," *Psychomusicology Music. Mind, Brain*, vol. 26, no. 3, pp. 257–269, 2016.

[8] A. Camurri, G. De Poli, A. Friberg, M. Leman, and G. Volpe, "The MEGA Project: Analysis and Synthesis of Multisensory Expressive Gesture in Performing Art Applications," *J. New Music Res.*, vol. 34, no. 1, pp. 5–21, 2005.

[9] C.-J. Tsay, "Sight over sound in the judgment of music performance." *Proc. Natl. Acad. Sci. USA*, vol. 110, no. 36, pp. 14 580–14 585, 2013.

[10] M. Rodger, C. Craig, and S. O'Modhrain, "Expertise is perceived from both sound and body movement in musical performance," *Hum. Mov. Sci.*, vol. 31, no. 5, 2012.

[11] H. F. Mitchell and R. a. R. MacDonald, "Listeners as spectators? Audio-visual integration improves music performer identification," *Psychol. Music*, vol. 42, pp. 112–127, 2012.

[12] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin, "Cross-modal interactions in the perception of musical performance." *Cognition*, vol. 101, no. 1, pp. 80–113, 2006.

[13] W. Thompson, P. Graham, and F. Russo, "Seeing music performance: Visual influences on perception and experience," *Semiotica*, vol. 156, pp. 177–202, 2005.

[14] M. Broughton and C. Stevens, "Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience," *Psychol. Music*, vol. 37, no. 2, pp. 137–153, 2009.

[15] C. J. Stevens, E. Schubert, R. H. Morris, M. Frear, J. Chen, S. Healey, C. Schoknecht, and S. Hansen, "Cognition and the temporal arts: Investigating audience response to dance using PDAs that record continuous data during live performance," *Int. J. Hum. Comput. Stud.*, vol. 67, no. 9, pp. 800–813, 2009.

[16] M. R. Thompson, M. M. Wanderley, and G. Luck, "Deadpan and immobile performance intentions share movement features but not expressive parameters," in *Proc. ICMPC and the 8th Conf. Eur. Soc. for Cogn. Sci. of Mus.*, Thessaloniki, 2012, pp. 987–988.

[17] S. Dahl and A. Friberg, "Visual perception of expressiveness in musicians' body movements," *Music Perc.*, vol. 24, pp. 433–454, 2007.

[18] W. Thompson, F. Russo, and L. Quinto, "Audio-visual integration of emotional cues in song," *Cogn. Emot.*, vol. 22, no. 8, pp. 1457–1470, 2008.

[19] A. Friberg, "A fuzzy analyzer of emotional expression in music performance and body motion," in *Proc. Music and Music Science*, Stockholm, 2004, pp. 1–13.

[20] J. Vuoskoski, M. Thompson, E. Clarke, and C. Spence, "Crossmodal interactions in the perception of expressivity in musical performance." *Atten. Percept. Psychophys.*, vol. 76, no. 2, pp. 591–604, 2014.

[21] J. Vuoskoski, M. Thompson, C. Spence, and E. Clarke, "Interaction of sight and sound in the perception and experience of musical performance," *Music Percept.*, vol. 33, no. 4, 2016.

[22] F. Upham and S. McAdams, "Activity Analysis and Coordination in Continuous Responses to Music," *Music Percept. an Interdiscip. J.*, vol. 35, no. 3, pp. 253–294, 2018.

[23] M. Febrero-Bande and M. O. de la Fuente, "Statistical computing in Functional Data Analysis: the R package fda.usc," *J. Stat. Softw.*, vol. 51, no. 4, 2012.

[24] J. A. Cuesta-Albertos and M. Febrero-Bande, "A simple multiway ANOVA for functional data," *Test*, vol. 19, pp. 537–557, 2010.

[25] M. O. Ernst and M. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, 2002.

[26] D. Alais and D. Burr, "The Ventriloquist Effect Results from Near-Optimal Bimodal Integration," *Curr. Biol.*, vol. 14, pp. 257–262, 2004.

[27] M. Ernst, "A Bayesian view on multimodal cue integration," in *Hum. Body Percept. from Inside Out*, G. Knoblich, I. Thornton, M. Grosjean, and M. Shiffar, Eds. New York: Oxford University Press, 2006, pp. 105–131.

[28] J. Hou, Y. Nam, W. Peng, and K. M. Lee, "Effects of screen size, viewing angle, and players' immersion tendencies on game experience," *Comput. Human Behav.*, vol. 28, pp. 617–623, 2012.

[29] J. Rigby, D. Brumby, A. Cox, and S. Gould, "Watching movies on netflix," in *Proc. 18th Int. Conf. Human-Comp. Interact. with Mob. Devices Serv. Adjun. - MobileHCI '16*. New York: ACM Press, 2016, pp. 714–721.

[30] S. O'Modhrain and R. B. Gillespie, *Once More, with Feeling: Revisiting the Role of Touch in Performer-Instrument Interaction*. Cham: Springer International Publishing, 2018, pp. 11–27.

[31] R. Dean and F. Bailes, "Modelling Perception of Structure and Affect in Music: Spectral Centroid and Wishart's Red Bird," *Empir. Musicol. Rev.*, vol. 6, no. 2, 2011.

[32] R. T. Dean, F. Bailes, and E. Schubert, "Acoustic intensity causes perceived changes in arousal levels in music: an experimental investigation," *PLoS One*, vol. 6, no. 1, pp. 1–8, 2011.

[33] J. Shim, L. G. Carlton, and J. Kim, "Estimation of Lifted Weight and Produced Effort through Perception of Point-Light Display," *Perception*, vol. 33, no. 3, pp. 277–291, 2004.

[34] M. Auvray, T. Hoellinger, S. Hanneton, and A. Roby-Brami, “Perceptual weight judgments when viewing one’s own and others’ movements under minimalist conditions of visual presentation,” *Perception*, vol. 40, pp. 1081–1103, 2011.

[35] J. Jolij and M. Meurs, “Music alters visual perception,” *PLoS One*, vol. 6, no. 4, 2011.