Article:

Emerson, Gina and Egermann, Hauke orcid.org/0000-0001-7014-7989 (2018) Gesture-sound causality from the audience's perspective: : investigating the aesthetic experience of performances with digital musical instruments. Psychology of Aesthetics, Creativity, and the Arts. pp. 96-109. ISSN 1931-3896

https://doi.org/10.1037/aca0000114

AUTHOR'S PREPRINT VERSION OF ACCEPTED PAPER
Psychology of Aesthetics, Creativity, and the Arts

Gesture-Sound Causality from the Audience's Perspective: Investigating the Aesthetic Experience of Performances with Digital Musical Instruments

Gina Emerson[1,2], and Hauke Egermann[2,3]

[1] Institute of Musicology, Humboldt-Universität zu Berlin, Germany
[2] Audio Communication Group, Technische Universität Berlin, Germany
[3] York Music Psychology Group, Department of Music, University of York, U.K.

Corresponding Author:

Hauke Egermann
York Music Psychology Group
Department of Music
University of York
Heslington, York
YO10 5DD, UK
hauke.egermann@york.ac.uk

Abstract

In contrast to their traditional, acoustic counterparts, digital musical instruments (DMIs) rarely feature a clear, causal relationship between the performer's actions and the sounds produced. They often function simply as systems for controlling digital sound synthesis, triggering computer-generated audio. This study aims to shed light on how the level of perceived causality of DMI designs impacts audience members' aesthetic responses to new DMIs. In a preliminary survey, 49 concert attendees listed adjectives that described their experience of a number of DMI performances. In a subsequent experiment, 31 participants rated video clips of performances with DMIs with causal and acausal mapping designs using the 8 most popular adjectives from the preliminary survey. The experimental stimuli were presented in their original version and in a manipulated version with a reduced level of gesture-sound causality. The manipulated version was created by placing the audio track of one section of the recording over the video track of a different section. It was predicted that the causal DMIs would be rated more positively, with the manipulation having a stronger effect on the ratings for the causal DMIs. Our results confirmed these hypotheses, and indicate that a lack of perceptible causality does have a negative impact on ratings of DMI performances. The acausal group received no significant difference in ratings between original and manipulated clips. We posit that this result arises from the greater understanding that clearer gesture-sound causality offers spectators. The implications of this result for DMI design and practice are discussed.

Keywords: Digital Music Instruments; Causality; Audience; Mapping; Multimodal

Gesture-Sound Causality from the Audience's Perspective: Investigating the Aesthetic Experience of Performances with Digital Musical Instruments

When experiencing a musical performance in a live concert setting, spectators receive auditory and visual information. On the basis of this, they determine which kinds of actions by the performer result in which kinds of sounds; in short, they perceive the live performance of cause-and-effect relationships. In the case of traditional acoustic instruments (e.g. the violin, piano or flute), this causal relationship is predictable, largely due to audience members' cultural familiarity with such instruments (Alperson, 2008; Caramiaux, Bevilacqua, Bianco, Schnell, Houix & Susini, 2014; Godøy, Haga & Jensenius, 2006). For performances with digital musical instruments (DMIs) however, defined here as systems designed to control digital sound synthesis (Wessel, Wright & Schott, 2002), conventional causalities between gesture and sound (e.g. that an action immediately results in a sound or that larger gestures result in louder sounds) are often absent. The present study reports empirical research carried out with the aim of exploring how this lack of a perceptible gesture-sound causality could impact spectators' aesthetic experience of performances with DMIs.[1]

Digital Musical Instruments and Causality

The absence of perceptible causality arises due to the physical separation between the sound control and sound generation elements in DMIs (Fig. 1). The process of DMI sound production begins with some form of input from a human performer, typically an action or a *gesture* (the latter is the standard term in DMI literature, e.g. Jensenius, 2014; Wessel et al., 2002) that is either executed on or with the control element of the instrument. The controller can take any form imaginable; existing trends include the replication or augmentation of the interfaces of existing acoustic instruments (e.g. Gurevich & von Muehlen, 2001; Impett, 1994; Machover, 1992; Schiesser & Schacher, 2012) and the development of 'wearable' controllers that are directly attached to the body (e.g. Nakra, 2000; Waisvisz, 1985; cf. Miranda & Wanderley, 2006 for further examples). The various input parameters (e.g. a striking or plucking action) are linked to their respective output parameter (a particular sound) via the *mapping*, the term for the configuration of the relationship between the control and sound generation parts of the instrument (Hunt & Kirk, 2000; Hunt & Wanderley, 2002). The final result, the actual audio output, is usually pre-coded in an audio programming environment on a laptop computer (e.g. Max/MSP, Csound, Pure Data, SuperCollider) and is triggered by the gestural input into the controller.

-place Figure 1 about here-

This form of sound production means in essence that DMIs need not have a stable, consistent way of producing sound. The mapping configuration is endlessly reprogrammable and no actual physical effort is required to create audio output from the system. For instrument designers and performers, this affords limitless flexibility for music-making and for making the performance of computer-based music more intuitive (Bovermann et al., 2014; Kim, 2012; Leman, 2008; Miranda & Wanderley 2006). In contrast, what DMIs offer audience members is less immediately clear. The performer's physical interaction with the device frequently does not appear to have a stable relationship to the sonic output, thus making it potentially difficult for spectators to discern how gestures are translated into sounds.

Several studies in the field of audiovisual perception research have explored the basis for the perception of causality between the visual and auditory streams, suggesting that it is largely the spatiotemporal correspondence of two events that enables humans to perceive one event as being caused by another (Spence, 2007). If a visible action seems to precede an appropriate sound, humans integrate the two streams of information and perceive the action as having caused the sound. The question under investigation here is whether causal stimuli are somehow aesthetically more pleasing to perceive than acausal stimuli. While there is little to no existing work directly on this topic, there are two strands of research that offer related insights, research on the general impact of visual information on the reception of musical performances and investigations into the reception of congruent or synchronous audiovisual stimuli.

Empirical Research on Audiovisual Music Perception

Various studies from this first line of research have demonstrated the fundamental principle that visual information accompanying the act of listening to music can influence what we hear and, at times, heighten our enjoyment of it (Behne & Wollner, 2011; for reviews, see Platz & Kopiez, 2012 and Schutz, 2008). Schutz and Lipscomb (2007), for example, report an illusion in which the duration of short-length tones played by a marimba player was perceived by participants as longer when accompanied by a video of the performer making a longer, larger performance gesture. Concerning the communication of musical tension through performance, Vines, Krumhansl, Wanderley & Levitin (2006) found that visual information can both increase and reduce perceived tension at different points in a piece, as revealed through an audio-only vs. visual-only vs. audiovisual paradigm involving ratings of performances of a Stravinsky clarinet piece. This was then extended to emotional responses; the intended emotion was shown to be most intensely perceived in the audiovisual condition (Vines, Krumhansl, Wanderley, Dalca & Levitin, 2011; see also Dahl & Friberg, 2007; Silveira, 2013; Vuoskoski, Thompson, Clarke, & Spence, 2013). These results indicate that visual input contributes significantly to the perception and appreciation of a musical performance, from which it can be inferred that DMI performances that present conflicting or confusing audiovisual information might be less effective at communicating musical intentions to an audience. Recent studies by Vuoskoski, Gatti, Spence and Clarke (2016) and Shoda and Adachi (2016) have provided contrasting results on this matter, reporting stronger emotional responses to audio-only conditions. However, these two studies both used music from the Romantic era (Brahms and Rachmaninoff respectively) that employ a musical idiom familiar to most listeners. The music performed typically on DMIs, at least those developed in academic settings, is very often atonal and in the form of unstructured improvisations, rather than fixed compositions, and therefore might require an audience member to rely more heavily on visual information when interpreting the expressive intent of a performance.

The second strand of relevant research centres around congruency theory, which relates to the human ability to judge the extent to which information from the auditory and visual streams occurs in synchrony (Cohen, 2013; Schutz & Kubovy, 2009). A number of investigations from the field of audiovisual multimedia research have suggested that congruency could be aesthetically pleasing to perceive. Iwamiya, Sugano and Kouda (2000) presented participants with short animated videos of a ball on a grid, which were accompanied by a simple audio track. For some clips, the downbeats of the audio track were temporally congruent with changes of perspective in the video and for others, this relationship was incongruent. The authors found that higher ratings of perceived temporal congruency correlated with higher impressiveness ratings of the clips. It can be suggested that congruent stimuli are aesthetically preferable because they are simpler to perceive; there is less conflicting information to be dealt with. This in turn implies that higher degrees of perceptible causality might similarly be aesthetically preferable. Unambiguous information could be expected to result in a higher level of understanding, which then informs aesthetic judgements (Juslin, 2013; Leder, Belke, Oeberst & Augustin, 2004).

DMI Reception Research

The majority of existing research directly related to DMIs has focused largely on the technical aspects of such devices, considering their attributes from the performer or instrument designer's perspective and thus evaluating them in terms of their potential expressivity and playability (O'Modhrain, 2011; Wanderley & Orio, 2002). More recently, a few papers regarding audience evaluations of DMIs have emerged. These have considered the perception of skill and the recognition of error in DMI performances (Fyans & Gurevich, 2011) and have explored audience understanding of various aspects of such systems, including mapping configurations (Barbosa, Calegario, Teichrieb, Ramalho & McGlynn, 2012) and audio output (Bergsland & Åse, 2012; Bown, Bell & Parkinson, 2014). Gurevich & Fyans (2011), for example, investigated spectators' perceptions of skill and error and interpretations of how the instruments functioned, concluding that many spectators had a poor understanding of relationship between performers' gestures and sound. The impact of this lack of understanding on audience members' *aesthetic experience* of the displayed DMIs was, however, not considered. Aesthetic responses to DMIs have been thematized in some theoretical contributions to the field; Croft (2007) proposes that a lack of perceptible gesture-sound causality can make performances with DMIs and similar technologies seem inauthentic and almost unrecognisable as live performances (see also Berthaut, Coyle, Moore & Limerick, 2015).

General Aims

The present study aims to address these open questions from both audiovisual causality studies and the existing work on DMI reception by investigating the aesthetic experience of causality in DMIs. To what extent does the level of perceived gesture-sound causality impact spectators' engagement with and appreciation of such performances? To achieve this aim, a two-part study was conducted, which comprised of a preliminary survey (Study 1) and an experiment (Study 2). The first study was carried out at a live concert to assess audience reactions to performances with new DMIs in an explorative manner and to collect common words and expressions that spectators use to describe their experiences of such a performance. We subsequently used these terms in the second experimental study in order to measure the aesthetic experience of DMI performances with varying degrees of causality.

Study 1: Preliminary Survey

Methods

Respondents

49 respondents completed the questionnaire during the concert (20 females, age in years: M = 29.8, range = 18-64). 30 respondents (61%) described themselves as being musically trained and 29 (59%) described themselves as being familiar with electroacoustic music.

Materials

The study was carried out at a concert, featuring four performances with new digital musical instruments developed in the context of a seminar at the University of XXXXXX, the partner institution of the XXXX project. Alongside *S/A/S/A,* an instrument that was also used in Study 2 (see Table 1 for details), these included a modular analogue synthesizer partly controlled via the movement of conductive concrete cubes on a plexiglass surface, an electronically-controlled percussion ensemble and a DMI with an organic controller made from a sheet of bacterial cellulose, a thermally conductive material stimulated during the performance through contact with human touch and breath.[2]

Procedure

The questionnaire featured two questions per instrument performance, the first of which asked the respondents to describe in note form what they had focused on during the performance (Question: 'What did you pay most attention to during the performance?'), while the latter requested a more general description of the respondents' views of the performance (Instruction: 'Describe the overall impression the performance made on you e.g. any emotional responses, thoughts on the instruments/music'). The questionnaires were handed to the audience members upon arrival at the concert and were filled out during the set changes between performances and during the interval. Responses could be given in either English or German.[3]

Results and Discussion

Question 1: Attention

The majority of answers to Question 1 mentioned having attended to the gestures or movements of the performers and their relationship with the resultant sounds. This result justifies the undertaking of further research into spectators' perceptions of this aspect of DMI performance. Answers included such terms as 'gesture' (n=7), 'movement' (n=21), 'relation' (n=8) and 'interaction' (n=16), with some respondents reporting having focused on interpreting the movements of specific parts of the body (e.g. the hands, n=25). Most notably, fourteen respondents specifically mentioned trying to understand the gesture-sound causality:

> I was trying to figure out how the moving of the cubes correspond to the changes in sound.
>
> (Respondent 16)

> [I was] trying to understand the logic of relation between [the] players, relating the blocks as causes to effects.
>
> (Respondent 9)

> [I focused on the] hands/hand movements and [their] connection with the sounds.
>
> (Respondent 21)

> How are the sounds produced?

(Respondent 20)

I wanted to understand how the sounds are produced, so I paid attention to the movements.

(Respondent 11)

It was a shame that there was a large gap between movement and interaction – no direct feedback [was] detectable.

(Respondent 49)

[It was] unclear for me, what was made [by the] laptop and what not.

(Respondent 30)

Such comments suggest that DMI mapping configurations can be the source of some confusion for audience members. Indeed, two respondents mentioned that a brief explanation of how the presented instrument works would have been helpful to them (Nos. 4 and 11). However, in contrast to this, one respondent (No. 10) stated that having to try to understand how the sound production worked added a sense of 'magic' to the performance.

Question 2: Overall Impressions

A summative content analysis was carried out on the answers to Question 2 (Hsieh & Shannon, 2005; Mayring, 2010). The words used to describe the performances were organized into three dimensions of spectator response that emerged from the data: terms pertaining to attention, emotional responses and words indicating preference or liking (Fig. 2). The most frequently used word overall was 'interesting' (n=27), followed by a number of other general, positive adjectives ('good', 'great' etc.) but answers also included more specific, emotionally-loaded terms such as 'tense'/'exciting' (n=13), 'threatening' (n=3) and ‘theatrical’/‘dramatic’ (n=4). Given that much of the music produced (perhaps with the exception of the percussion setup) was freely structured, atonal and often involved only abstract, isolated sounds, it is notable that such a wide range of adjectives was employed and that emotional responses were reported; prior research has indicated that atonal music provokes weaker emotional responses in listeners than tonal music (Daynes, 2011). This may well be due to the respondents’ high overall level of familiarity with electroacoustic/experimental music.

-place Figure 2 about here-

Overall, our preliminary study confirmed that the perception of gesture-sound causality constitutes an important part of the audience experience of a concert with new DMIs. This insight and the collected terms from Question 2 served as a useful basis for the design of the experiment.

Study 2: Experiment

Aims and Hypotheses

The aim of this study was to investigate the aesthetic experience of DMI performances with varying degrees of causality. We asked participants to rate video clips of performances for two groups of DMIs: one with mapping designs that were more likely to be perceived as *causal* and one with mapping designs that were more likely to be perceived as *acausal*. These clips were presented in their original version and in a manipulated version, a mismatch of the audio and video tracks that created a unified acausal comparison condition. In light of the existing research detailed above, which posits connections between (a) causality and higher levels of audience understanding and (b) between congruent stimuli, attention and effectiveness, we made the following hypotheses:

Perceived Causality H1:

a) The DMIs in the causal group will be perceived as more causal than the acausal group.

b) The original recordings will be rated as more causal than the manipulations.

c) There will be a significant interaction between the factors Instrument Category and Manipulation, such that the effect of the manipulation on perceived causality ratings will be stronger for causal DMIs than for acausal DMIs.

Overall Experience H2:

a) The overall experience of DMIs in the causal group will be rated more positively (i.e. as more interesting/enjoyable) than the acausal group.
b) The original recordings will be rated more positively than the manipulations.
c) There will be a significant interaction between the factors Instrument Category and Manipulation, such that the effect of the manipulation on aesthetic experience ratings will be stronger for causal DMIs than for acausal DMIs.

Method

Participants

Thirty-one participants (17 females, age in years: M = 26.8, range = 20-51, SD = 5.53) took part in the experiment, after having been recruited via university mailing lists. Four were graduate students at the XXX Group of the Technical University, XXXX participating for credit points. The remainder were other students from a variety of programmes, who received 10 Euros for participating. None of the participants had attended the concert at which preliminary study was conducted. The participants had received an average of 4.9 years of musical training (range = 0-27 years, SD = 5.66 years), with eighteen self-reporting as amateur musicians, twelve as non-musicians and one as a professional. Participants were asked to rate on scales of one to seven their answers to the questions, ‘How often do you listen to experimental music?’ and ‘How much do you like experimental music?’. The results suggest that although the sample generally had quite a low exposure to experimental music, their preference for it was relatively high (M = 2.77, SD = 1.77 and M = 4.19, SD = 1.6 respectively).

Materials

Five DMIs, which had all, except one, been developed within the XXX project, were recorded to create the audiovisual stimuli for the experiment, using a JVC GY-DV301 camcorder with direct audio input from the instruments. When selecting the instruments for the recording, we took care to gather a wide range of controllers with differing degrees of perceptible causality. Three of the five controllers were then categorized as being more causal than the other three, resulting in two groups of instruments. This classification centred around the perceptibility of the mapping design and was based on a total of four factors that draw on Croft's conditions for liveness and instrumentality in DMIs (Croft 2007, p. 64-65) and on the five spectator questions used by Gurevich and Fyans (2011):

1. **Type of controller:** DMIs that make use of interfaces resembling existing acoustic instruments are more likely to facilitate the understanding of gesture-sound relationships via the familiarity they offer; interfaces that allow for greater gestural variation from the performer (e.g. a wearable controller vs. a more laptop-based setup) could also have a higher degree of perceptible causality.
2. **Latency/Temporal Delay:** while this can be difficult to avoid, a significant amount of latency is likely to impact the perceptibility of the mapping design and overall believability of the instrument, this could also be the case for other forms of temporal delay (e.g. delay/reverb effects or looping in the sonic output).
3. **Consistency of mapping design:** the relationship between the performer's actions and the resulting sounds should be consistent to aid perceptibility.
4. **Resulting output is appropriate for the given input:** larger gestures should map onto increased volume; short or percussive sounds should result from

short, discrete actions; no movement should mean no sound, continuous sound should involve some kind of continuous effort.

The categorisations were made on the basis of how the instrument appeared in the recording (see Table 1 for the instrument details); since DMIs are endlessly reprogrammable, it is difficult to make a decisive classification. *PushPull*, for instance, features a mode that randomizes the mapping strategy at will (a potentially ‘acausal’ feature from the audience’s perspective) but this mode was not employed during the recording.

The performers each played a short performance of approximately 5-10 minutes in length. A possible structure for the musical content was suggested to the performers, which proposed the inclusion of contrasting material (e.g. a section of abstract sounds followed by more tonal, melodic passages). Ideally, the music played by each instrument would have been kept the same (the performers could have been given a composition to perform) but given the vast differences in timbre and playing style between the instruments, uniformity in output could simply not have been achieved. The request to play in different styles resulted in one to three different performance styles or ‘modes’ per recording. In order to produce a unified acausal condition for all instruments, the videos were manipulated in *VideoPad* Version 4.07 as follows. A 1-minute clip was selected from each of the 10 modes available. In all cases, this corresponded to 00:30 to 01:30 of the recording; the first 30 seconds were always ignored so that the clips would show the performances at a more developed stage. For all of the instruments with two available modes, the audio from the second 1 minute clip was placed over the video of the first (Mode 2 Audio / Mode 1 Video) and vice versa (Mode 1 Audio / Mode 2 Video), resulting in two original and two manipulated versions.

--- place Table 1 about here ---

For *S/A/S/A* (three modes), the audio from Mode 2 was placed over the video from

Mode 1, the audio from Mode 3 over Mode 2's video and the audio from Mode 1 over Mode 3's video, creating a total of six clips. Finally, for *The Finger* (one mode), a 2 minute clip was selected, cut in half and the audio from the latter half placed over the first half and vice versa. This resulted in a total of 22 one minute clips, eleven originals and eleven manipulations (summarized in Table 1 above). The clips (video: 1280x720 pixel, ca. 3.5 Mbit/s, h264 format; audio: 2 channel mp3, 192 kbit/s) were presented with *PsychoPy* Version 1.82.01 (Peirce, 2016), running on a Lenovo Z500 series laptop, which was connected to an external 24" LCD monitor and Philips SPA 5300 2.1 desktop loudspeakers. The same computer and software were used for data collection. The participants could adjust the volume as desired.

Procedure

The participants each watched all 22 clips once. The order that the clips were played in was randomized for each participant. After each clip, they were asked to rate their level of agreement with the following adjectives on a scale from one to seven: *interesting, good, exciting, pleasant, boring, repetitive/predictable, stimulating, calming, virtuosic*. With the exception of 'virtuosic', these were the terms that occurred most frequently in the preliminary study data (see Fig. 2). Terms with very similar meanings were excluded e.g. 'good' was used to cover 'nice', 'excellent' and 'cool', even though these were frequently used in their own right. The rating of virtuosity was included because it has been a common topic in DMI research and is thought to be a possible factor in the evaluation of such devices by spectators (Gurevich & Fyans 2011). It has not, however, previously been looked at in a quantitative study or specifically in relation to gesture-sound causality perception. Two further questions followed the adjective ratings, one asking how much attention the participant paid to the performance (Question: 'How fixating was the performance?') and in how far the participant thought that the music was influenced by the gestures of the performer (Question: 'In your opinion, to what extent was the music influenced by the performer's movements?'). At the very end, a demographic questionnaire was filled out, which also asked participants whether more information about the instruments would have helped their evaluation. The participants were kept naive to the manipulation and were not given any information about any of the instruments either before or during their participation in the experiment.

Results and Discussion

To simplify the data for further analysis, a principal component analysis (PCA) was carried out on the response variables. The items 'Perceived Causality' and 'Virtuosity' were excluded from the PCA as they are distinct from the other ratings and do not fit into the emotion/attention response model. A correlation matrix for the response variables was produced and it was found that the item 'Repetitive/Predictable' correlated very poorly with almost all other response variables (only one coefficient > 0.3). It was thus excluded from the analysis.

The PCA was conducted on eight items with orthogonal rotation (varimax). The Kaiser-Meyer-Olkin (KMO) statistic confirmed the sampling adequacy for the analysis (KMO = 0.903; all KMO values for individual items were > 0.6) and Bartlett's test of sphericity identified correlations of sufficient size between items to justify the PCA (X^2 (28) = 4682.68, $p < 0.01$). The initial analysis revealed two components with eigenvalues over 1, which together explained 81.9% of the variance. The two components were confirmed by a scree plot based on the elbow-criterion and were therefore retained for the rotation. Table 2 shows the component loadings after rotation. The loadings suggest that Component 1 can be interpreted in terms of the arousal-valence model as 'High Arousal, Positive Valence' and Component 2 as 'Low Arousal, Positive Valence' (see Russell 1980). Component scores were extracted using the Anderson-Rubin method and used for the calculation of further analyses. The data was reduced further by averaging the scores for the two manipulated and two original conditions for each instrument and then calculating mean scores for the causal and acausal groups and their manipulations.

--- place Table 2 about here ---

Perceived Causality

In order to test the effect of the manipulation and of instrument group on perceived causality and thereby confirm whether or not the stimuli were received as expected (Hypotheses 1a-c), we conducted a two-way repeated measures ANOVA (Manipulation and Instrument Category). The results show that the manipulation had a significant effect on perceived causality ratings, $F(1, 30) = 33.11$, $p < .01$, $\eta^2 = 0.72$ and that there was furthermore a significant difference between the instrument groups in the results for this rating, $F(1, 30) = 19.32$, $p < .01$, $\eta^2 = 0.12$. The interaction effect was significant, implying that the effect of manipulation on perceived causality varied significantly according to instrument category, $F(1, 30) = 4.51$, $p < .05$, $\eta^2 = 0.02$.

A set of four post hoc paired sample t-tests was conducted to explore the interaction effect in more depth (see Fig. 5a and Appendix Table 1). The participants' ratings of perceived causality confirmed the categorisation of the instruments (Hypothesis 1a) and the causality ratings were significantly lower in the manipulated condition, indicating that the experimental manipulation was accurately produced (Hypothesis 1b). The manipulation also had a stronger effect on ratings of perceived causality for the causal DMIs than for the acausal DMIs (Hypothesis 1c).

--- place Figure 5 about here ---

Component 1: High Arousal, Positive Valence

A two-way repeated measures ANOVA (Manipulation and Instrument Category) was conducted to test the effect of the manipulation and of instrument group on the positivity of ratings for Component 1 (Hypotheses 2a-c). This was calculated using the component scores from the PCA. The results show that the manipulation had a significant effect on the rating items that comprise Component 1 (Exciting, Interesting, Attention Paid, Boring, Stimulating and Good), $F(1, 30) = 15.93$, $p <.01$, $\eta^2 = 0.2$, and that there was furthermore a significant difference between the instrument categories in the results for this component,

$F(1, 30) = 59.10$, $p < .01$, $\eta^2 = 0.63$. The interaction effect was also significant, suggesting that the effect of the manipulation on the ratings of the Component 1 items varied significantly between the two groups, $F(1, 30) = 18.75$, $p < .01$, $\eta^2 = 0.06$.

The results of the post hoc t-tests (see Fig. 5b and Appendix Table 2) show that the causal group was indeed rated as significantly more positively stimulating than the acausal instruments (Hypothesis 2a) and that the manipulations were in general rated as less positively stimulating than the originals (Hypothesis 2b). The ratings for the causal group showed a significant difference between original and manipulated versions but this was not the case for the acausal group, indicating that they were far less strongly affected by the manipulation (Hypothesis 2c).

Component 2: Low Arousal, Positive Valence

A further two-way repeated measures ANOVA (Manipulation and Instrument Category) was conducted to test the effect of the manipulation and instrument group on ratings for Component 2 (Hypotheses 2a-c). Once again, this was calculated using the component scores from the PCA. The manipulation had a significant effect on the rating items that comprise Component 2 (Calming, Pleasant), $F(1, 30) = 8.89$, $p <.01$, $\eta^2 = 0.04$ and there was furthermore a significant difference between the two groups in the results for this component, $F(1, 30) = 45.94$, $p <.01$, $\eta^2 = 0.84$. The interaction effect was not significant, $F(1, 30) = 2.74$, $p = 0.108$, $\eta^2 = 0.01$, but post hoc t-tests (see Fig. 5c and Appendix Table 3) still revealed a significant difference in Component 2 ratings between the original and manipulated conditions for the causal group and not for the acausal group (Hypothesis 2c; the differences between the versions, independent of the instrument category component, also support Hypothesis 2b). The acausal group, however, received significantly higher ratings than the causal group on Component 2, in contrast to Hypothesis 2a; this result is interpreted in the General Discussion below.

Virtuosity

A final two-way repeated measures ANOVA (Manipulation and Instrument Category) was conducted to test the effect of the manipulation on ratings of virtuosity (Hypotheses 2a-c). The results indicate that the manipulation had a significant effect on ratings of virtuosity, $F(1, 30) = 21.24$, $p < .01$, $\eta^2 = 0.48$, and that there was a significant difference between the instrument groups in the results for this response variable, $F(1, 30) = 13.39$, $p < .01$, $\eta^2 = 0.25$. The interaction effect was also significant, implying that the effect of manipulation on the virtuosity ratings varied significantly between the causal and acausal groups, $F(1, 30) = 15.26$, $p <.01$, $\eta^2 = 0.09$.

The results of the post hoc t-tests (see Fig. 5d and Appendix Table 4) revealed significantly lower ratings of virtuosity for the acausal group in comparison to the causal instruments (Hypothesis 2a) and lower ratings of virtuosity for the manipulations than for the originals (Hypothesis 2b). This difference was only just significant for the Acausal Original vs. Acausal Manipulated pair, $p = 0.012$, in contrast to the large significant difference for the causal group, which shows once more that the manipulation had a stronger effect on ratings for the causal DMIs (Hypothesis 2c).

General Discussion

The reported study aimed to investigate the aesthetic experience of DMI performances with varying degrees of causality. Overall, our results indicate that a lack of perceptible causality does have a negative impact on ratings of performances with DMIs. The performances on instruments in the causal group, *PushPull*, *The Finger* and *S/A/S/A*, were rated on average as considerably more interesting, and, more successful at holding the participants' attention (Component 1) and as demonstrating more skill than the acausal instruments and in comparison to their manipulated versions (Virtuosity ratings). We posit that this result arises from the greater understanding that clearer gesture-sound causality offers spectators. The ability to perceive reliable relationships between gesture and sound establishes a necessary foundation for further judgements of a performance, as summarized in our proposed model of the perception and judgement process (Fig. 6; cf. BRECVEMA model in Juslin, 2013). This notion is supported further by the finding that the manipulations had, as predicted, a much stronger impact on the ratings for the causal DMIs than on those for the acausal DMIs.

--- place Figure 6 about here ---

The acausal group received no significant difference in ratings between original and manipulated versions for perceived causality ratings, Component 1, and Component 2. This suggests that participants perhaps struggled to perceive much of a difference between the original and manipulated versions of the acausal DMIs, which in turn implies that they were less able to figure out what the gesture-sound relationship was supposed to be and had thereby a poorer understanding. If a manipulation creates almost the same impression upon an audience as the original performance, this also shows how a lack of perceptible gesture-sound causality could thwart performers' attempts at communicating an artistic or expressive goal to spectators. There was, however, a significant difference in the Virtuosity ratings for the original and manipulated versions of the acausal clips; the participants did judge there to be less skill displayed in the manipulated versions. The impact of the manipulation on Virtuosity ratings was, as predicted, still much stronger for the causal instruments.

The results for the Low Arousal Component are, at first glance, more difficult to interpret; given that the acausal instruments were expected to evoke a negative response, their high rating on the Low Arousal, Positive Valence component (which includes the items 'Calming' and 'Pleasant') may seem to contradict Hypothesis 2a. However, the results require interpretation in combination with those from Component 1 which is summarized in Fig. 7. Here, the overall mean ratings of the four instrument conditions are projected onto a two-dimensional emotional space (subsequently rotated through 90 degrees to match the valence/arousal model by Russell, 1980). Instruments in the causal group were perceived as stimulating/arousing and positive, and the audiovisual manipulation provoked a shift in ratings to the negative side of the emotion space, supporting Hypothesis 2b. However, instruments in the acausal group were in general perceived as relatively calming, but with a mean valence rating close the middle of the space, indicating that they were perceived as boring or neutral. Furthermore, the audiovisual manipulation did not change this position, as there were no significant differences between the original and manipulated versions of the acausal instruments for ratings of Component 1 and 2.

--- place Figure 7 about here ---

The general difference in arousal between the causal and acausal groups could have been confounded by the music played on the instruments in the acausal group, which could well have been perceived as calming independent of the instrument design. To tackle the possible confounding factors in the study, an alternative design could have involved creating DMIs specifically for the experiment. This would have potentially allowed us to control a number of further factors, including the music played on the instruments, the appearance of different performers in the video clips and greater control over the differing degrees of gesture-sound causality displayed. Gurevich and Fyans (2011) and Berthaut et al. (2015), for instance, both employ instruments designed specifically for the experiment (see also Marquez-Borbon, Gurevich, Fyans & Stapleton, 2011). While it is certainly appealing to attempt to control these further aspects, such paradigms ultimately lose a lot in terms of ecological validity. By using recordings of a range of DMIs that were already in existence, the study presented here can be considered to well reflect the reality of DMI practice.

The result of our study carries potentially controversial implications for DMI practice. If a lack of perceptible gesture-sound causality does negatively impact spectators' aesthetic experiences of DMI performances, how should DMI designers address this issue? One approach would be to draw directly on the findings presented here and design DMIs that allow for a more clearly perceptible gesture-sound causality. For the instruments in the causal experimental group, this perceptibility manifested itself primarily in the amount of visual information the device allowed for and mappings that were logical from the audience's perspective. Indeed, the factors impacting mapping perceptibility (type of controller, latency or other temporal delay, consistency of mapping design and having an appropriate input-output response) could be taken up as guidelines for future DMI designs (see similar propositions in Barbosa et al., 2012; Croft, 2007 and O'Modhrain, 2011). For many, however, delineating a framework for DMI design such as this could represent too much of an imposition on creative practice.

An alternative approach would be to consider how to make mappings more transparent to audiences in terms of performance presentation. Directly offering information to audiences on the mapping design and how the instrument produces sound would be a solution that audiences would likely welcome; in response to the final question of whether more information about the instruments would have helped them evaluate the performances, 58% (n=18) of participants from the second study responded affirmatively. This information could take the form of programme notes or pre-performance demonstrations but there are some proposed methods for providing clarification during a performance. Berthaut, Marshall, Subramanian and Hachet (2013) have, for example, designed a display system that can be placed under tabletop DMIs, which illustrates, via 3D visualisations, the ongoing sound generation process (a version of this was tested in Berthaut et al. 2015). There are also several live coding performers who include projections of their screen display in their setup, allowing spectators to see that changes in the code affect the sounds generated, from which a basic sense of the gesture-sound causality can be extracted (Brown & Sorensen, 2007; 2009; McLean, Griffiths, Collins & Wiggins, 2010).

These various strategies could lead into directions for further research. One possible line of investigation could seek to establish the extent to which audiences understand or receive artistic intention accurately and could involve both the audience's and the designer/performer's perspective in this. Artists could be asked to note down their intentions and their means for realising them (e.g. the expression of particular emotion or the creation of a certain atmosphere) prior to performance and then the extent to which this matches the expression as perceived by the audience could be qualitatively evaluated for a number of different DMIs. A similar quantitative paradigm could investigate the communication of expressive intent via the comparison of audiovisual and video-only conditions. Through this, it could be established whether or not DMI performances provide enough visual information for the recognition of emotional expression without audio, which has shown to be possible for performances with acoustic instruments (Vines et al. 2006; Vines et al. 2011). The potential for supplementary information to modulate the accuracy of musical communication could be added to these paradigms and could also thereby compare the efficacy of different modes of providing information to audiences (e.g. programme notes vs. pre-performance demonstration vs. a visual presentation accompanying the performance).

Conclusions

The studies presented here show that the relationship between gesture and sound is an important factor for audiences attending DMI performances. A higher degree of perceptible causality, as created by the mapping and the type of controller, provides spectators with more information and more reference points for evaluating the performance. Being able to perceive, understand, and then create a mental model of the sound generation process not only generates greater interest, it also appears to provide a basis for assessing the amount of skill displayed in the performance.

This result could be of use to those involved in creating new DMIs. Designers should aim to produce richer musical and perceptual experiences that engage and provoke audiences without alienating them. This is essential in order to move away from instrument design for its own sake and towards the production of new digital musical instruments that are not only interesting to play, but also to perceive.

References

Alperson, P. (2008). The Instrumentality of Music. *The Journal of Aesthetics and Art Criticism*, *66*(1), 37–51. doi: 10.1111/j.1540-594X.2008.00286.x

Barbosa, J., Calegario, F., Teichrieb, V., Ramalho, G., & McGlynn, P. (2012). Considering Audiences' View Towards an Evaluation Methodology for Digital Musical Instruments. *NIME 2012: Proceedings of the 12th International Conference on New Interfaces for Musical Expression*, Ann Arbor, MI (pp. 403–408). Retrieved from http://www.nime.org/proceedings/2012/nime2012_174.pdf

Behne, K.-E., & Wollner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae*, *15*, 324–342. doi: 10.1177/1029864911410955

Bergsland, A., & Åse, T. (2012). Using a seeing / blindfolded paradigm to study audience experiences of live-electronic performances with voice. *NIME 2012: Proceedings of the 12th International Conference on New Interfaces for Musical Expression*, Ann Arbor, MI (pp. 433–436). Retrieved from: http://www.nime.org/proceedings/2012/nime2012_168.pdf

Berthaut, F., Coyle D., Moore, J. W., Limerick H. (2015). Liveness Through the Lens of Agency and Causality. *NIME 2015: Proceedings of Conference on New Interfaces for Musical Expression,* Baton Rouge, LA (pp. 382-6). Retrieved from http://www.nime.org/proceedings/2015/nime2015_272.pdf

Berthaut, F., Marshall M., Subramanian S., Hachet M. (2013). *Rouages*: Revealing the Mechanisms of Digital Musical Instruments to the Audience. *NIME 2013: Proceedings of the International Conference on New Interfaces for Musical Expression*, Daejeon, South Korea (pp. 165–169). Retrieved from http://nime.org/proceedings/2013/nime2013_51.pdf

Bovermann, T., Egermann, H., Foerstel, A., Hinrichsen, A., Hildebrand Marques Lopes, D., Pysiewicz, A., Hardjowirogo, S., Weinzierl, S., & de Campo, A. (2014). 3DMIN – Challenges and Interventions in Design, Development and Dissemination of New Musical Instruments. *Proceedings of ICMC/SMC*, Athens. Retrieved from http://www.3dmin.org/wp-content/uploads/2014/10/2014-icmc_intro3dmin-final.pdf

Bown, O., Bell, R., & Parkinson, A. (2014). Examining the Perception of Liveness and Activity in Laptop Music: Listeners' Inference about what the Performer is Doing from the Audio Alone. *NIME 2014: Proceedings of the 14th International Conference on New Interfaces for Musical Expression*, London, U.K. (pp. 13–18). Retrieved from http://www.nime.org/proceedings/2014/nime2014_538.pdf

Brown, A. R. & Sorensen, A. C., (2007). aa-cell in Practice: An Approach to Musical Live Coding. *Proceedings of the International Computer Music Conference*, 292-299. Retrieved from http://eprints.qut.edu.au/39768/1/c39768.pdf

Brown, A. R., & Sorensen, A. C. (2009). Interacting with Generative Music through Live Coding. *Contemporary Music Review*, *28*(1), 17–29. doi: 10.1080/07494460802663991

Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O., & Susini, P. (2014). The Role of Sound Source Perception in Gestural Sound Description. *ACM Transactions on Applied Perception*, *11*(1), 1–19. doi: 10.1145/2536811

Cohen, A. (2013): Congruence-Association Model of Music and Multimedia: Origin and Evolution. In Tan, S. Cohen A. J., Lipscomb S. D. & Kendall, R. (Eds.), *The Psychology of Music in Multimedia* (pp. 17-138). Oxford: Oxford University Press.

Croft, J. (2007). Theses on Liveness. *Organised Sound*, *12*(01), 59-66.

Dahl, S. & Friberg, A. (2007): Visual Perception of Expressiveness in Musicians' Body Movements. *Music Perception: An Interdisciplinary Journal. 25*(5): 433–54. doi: 10.1525/mp.2007.24.5.433

Daynes, H. (2011). Listeners' perceptual and emotional responses to tonal and atonal music. *Psychology of Music*, *39*(4), 468–502. doi: 10.1177/0305735610378182

Fyans, A.C. & Gurevich M. (2011). Perceptions of Skill in Performances with Acoustic and Electronic Instruments. *NIME 2011*: *Proceedings of the 11th International Conference on New Interfaces for Musical Expression,* Oslo, Norway (pp. 495–498). Retrieved from http://www.nime.org/proceedings/2011/nime2011_495.pdf

Godøy, R. I., Haga, E., & Jensenius, A. R. (2006). Playing "air instruments": mimicry of sound-producing gestures by novices and experts. In Gibet, S., Courty N., Kamp J.-F. (Eds.) *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers* (pp. 256-67). Berlin, Heidelberg: Springer.

Gurevich, M. & Fyans, A.C. (2011). Digital Musical Interactions: Performer–system relationships and their perception by spectators. *Organised Sound*. 16(2): 166-75. doi: 10.1017/S1355771811000112

Gurevich, M., & von Muehlen, S. (2001). The Accordiatron: A MIDI controller for interactive music. *NIME 2001: Proceedings of the 1st International Conference on New Interfaces for Musical Expression,* Seattle, WA (pp. 1-3). Retrieved from http://www.nime.org/proceedings/2001/nime2001_027.pdf

Hinrichsen, A., Hardjowirogo, S., Hildebrand Marques Lopes, D., Bovermann, T. (2014). *PushPull*. Reflections on Building a Musical Instrument Prototype. *Proceedings of the 2nd International Conference on Life Interfaces,* Lisbon, Portugal (pp. 196-207). Retrieved from http://www.3dmin.org/wp-content/uploads/2014/03/Hinrichsen_2014a.pdf

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. doi: 10.1177/1049732305276687

Hunt, A., & Kirk, R. (2000). Mapping strategies for musical performance. In Wanderley, M. & Battier, M. (Eds.) *Trends in Gestural Control of Music* (pp. 231-258). Paris: IRCAM.

Hunt, A., & Wanderley, M. (2002). Mapping performer parameters to synthesis engines, *Organised Sound. 7(2),* 97–108. doi: 10.1017/S1355771802002030

Impett, J. (1994). A meta-trumpet(-er). *Proceedings of the 20th International Computer Music Conference*, Aarhus, Denmark (pp. 147-150). Retrieved from http://quod.lib.umich.edu/cgi/p/pod/dod-idx/meta-trumpeter.pdf?c=icmc;idno=bbp2372.1994.037

Iwamiya, S., Sugano, Y., & Kouda, K. (2000): The Effects of Synchronization of Temporal Structure of Sound and Motion Picture on the Impression of Audiovisual Context. *IEEE 12th International Conference on Systems, Man & Cybernetics*, Nashville, TN (pp. 1222-1225).

Jensensius, A.R. (2014). To Gesture or Not? An Analysis of Terminology in NIME Proceedings 2001-2013. *NIME 2014: Proceedings of the 14th International Conference on New Interfaces for Musical Expression*, London, U.K. (pp. 217–220). Retrieved from http://www.nime.org/proceedings/2014/nime2014_351.pdf

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235-266. doi: 10.1016/j.plrev.2013.05.008

Kim, J. H. (2012). *Embodiment in interaktiven Musik- und Medienperformances – unter besonderer Berücksichtigung medientheoretischer und kognitionswissenschaftlicher Perspektiven*. Osnabrück: epOs.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489-508. doi: 10.1348/0007126042369811

Lerdahl, F. (1992). Cognitive constraints on compositional systems. *Contemporary Music Review*, *6*(2), 97–121. doi: 10.1080/07494469200640161

Lerdahl, F. (2001). *Tonal Pitch Space*. Oxford: Oxford University Press.

Leman, M. (2008). *Embodied Music Cognition and Mediation Technology*. Cambridge, London: MIT Press.

Machover, T. (1992). *Hyperinstruments*: *A Progress Report*. MIT Media Lab.

Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim and Basel: Beltz Verlag.

Marquez-Borbon, A., Gurevich, M., Fyans, A., & Stapleton, P. (2011). Designing digital musical interactions in experimental contexts. *Proceedings of the 11th International Conference on New Interfaces for Musical Expression,* Oslo, Norway, (pp. 373-376). Retrieved from http://www.nime.org/proceedings/2011/nime2011_373.pdf

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*. 264: 746–748. doi: 10.1038/264746a0

McLean, A., Griffiths, D., Collins, N. & Wiggins G. (2010). Visualisation of live code. In SealA., Bowen, J. & Ng. K. (Eds.) *EVA London 2010: Electronic Visualisation and the Arts*, (pp. 26–30). London: BCS.

Miranda, E. R. & Wanderley M.M. (2006): *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*. Computer Music and Digital Audio Series, Vol. 21. Middleton: AR Editions.

Nakra, T. M. (2000). Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture. Unpublished PhD Thesis. MIT Media Lab.

O'Modhrain, S. (2011). A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal*, *35*(1), doi: 28–42. 10.1162/COMJ_a_00038

Peirce, J. (2016). *PsychoPy: Psychology Software for Python*. Retrieved from http://www.psychopy.org/PsychoPyManual.pdf

Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation en- hances the appreciation of music performance. *Music Perception 30*(1)*,* 71-83. doi: 10.1525/MP.2012.30.1.71

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-78. doi: http://dx.doi.org/10.1037/h0077714

Schiesser, S., & Schacher, J. C. (2012). SABRe : The Augmented Bass Clarinet. *NIME 2012: Proceedings of the 12th International Conference on New Interfaces for Musical Expression*, Ann Arbor, MI (pp. 109–112). Retrieved from http://www.nime.org/proceedings/2012/nime2012_193.pdf

Schutz, M. (2008). Seeing Music? What musicians need to know about vision. *Empirical Musicology Review*, *3*(3), 83–108. Retrieved from https://kb.osu.edu/dspace/bitstream/handle/1811/34098/EMR000053a_revised.pdf

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(6), 1791–1810.

Schutz, M., & Lipscomb, S.D. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*(6), 888–897. doi: 10.1068/p5635

Shoda, H., & Adachi, M. (2016). Expressivity, Affective Nuance, and Presentation Modality in a Performer-to-Audience Communication. *Psychomusicology: Music, Mind and Brain, 26*(2), 167-178. doi: 10.1037/pmu0000141

Silveira, J. M. (2013). The effect of body movement on listeners' perceptions of musicality in trombone quartet performance. *International Journal of Music Education*. doi: 10.1177/0255761413491210

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61–70. doi: 10.1250/ast.28.61

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*(2), 157–170. doi: 10.1016/j.cognition.2010.11.010

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80–113. doi: 10.1016/j.cognition.2005.09.003

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2013). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception & Psychophysics*. 76: 591–604. doi: 10.3758/s13414-013-0582-2

Vuoskoski, J. K., Gatti, E., Spence, C., & Clarke, E. F. (2016). Do Visual Cues Intensify the Emotional Responses Evoked by Musical Performance? A Psychophysiological Investigation. *Psychomusicology: Music, Mind, and Brain, 26*(2), 179-188. doi: 10.1037/pmu0000142

Waisvisz, M. (1985) *The Hands*: A Set of Remote MIDI-Controllers. *Proceedings of the 10th International Computer Music Conference*, Burnaby, Canada (pp. 86–89). Retrieved from http://quod.lib.umich.edu/i/icmc/bbp2372.1985.049/1

Wanderley, M. M.; Orio, N. (2002). Musical Expression: Borrowing Tools. *Computer Music Journal*, *26*(3), 62–76. Retrieved from http://www.jstor.org/stable/3681979

Wessel, D., Wright, M., & Schott, J. (2002). Intimate Musical Control of Computers with a Variety of Controllers and Gesture Mapping Metaphors. *NIME 2002: Proceedings of the 2nd International Conference on New Interfaces for Musical Expression,* Dublin, Ireland (pp. 1–3). Retrieved from http://www.nime.org/proceedings/2002/nime2002_192.pdf

Footnotes

[1] 'Aesthetic experience' is used here to refer to a spectator's general affective experience and enjoyment of an artistic performance (see Juslin, 2013).

[2] Video clips of the instruments can be found here: http://www.3dmin.org/activity/3dmin-concert-series/ws1415-student-concert/, last accessed 27.10.2015

[3] We use English equivalents for German words when discussing the results. Equivalent German terms are included in the scores given for word usage frequency.

Acknowledgements

We would like to thank all participants and musicians for their contribution to this study. Furthermore, we are highly grateful for the contribution of XXXX in designing the experiment and producing the stimuli used.

Table 1

Details of the Five DMIs Recorded for the Experiment

Name	Creator(s)	Description	Causal or Acausal?	No. of Performance Modes	Manipulation Method
The Bass *(Fig. 4)*	Pierre-Alexandre Tremblay	**Augmented instrument** **Hardware:** 6-string fretless bass guitar played with distortion and volume pedals **Software:** Feeds via RME 400 sound card to laptop running Max/MSP performance patches, further control via iPad running Lemur	**Acausal** *Despite familiar interface, effects and looping create unclear gesture-sound relationship via temporal delay*	2	Mode 2 Audio / Mode 1 Video, Mode 1 Audio / Mode 2 Video
The Finger *(Fig. 3)*	Dominik Hildebrand Marques-Lopez	**Wearable controller** **Hardware:** x-OSC I/O board combined with JeeNode wireless board and GravityPlug accelerometer, 2 potentiometers and 3 push buttons for control, plexiglass for outer shell of the glove **Software:** SuperCollider	**Causal** *Height and speed of gestures map logically onto changes in pitch/volume, no movement = no sound*	1	2 min clip selected, halved, 2nd Half Audio / 1st Half Video, 1st Half Audio / 2nd Half Video
Jerry *(Fig. 4)*	Christoph Schultz, Marten Seedorf	**Laptop instrument** **Hardware:** LogiLink USB Number Keypad and optical mouse **Software:** Pure Data	**Acausal** *Laptop-based setup provides little visual information, continuous sound produced without continuous visible effort*	2	Mode 2 Audio / Mode 1 Video, Mode 1 Audio / Mode 2 Video
PushPull *(Fig. 3)*	Till Bovermann, Dominik Hildebrand Marques-Lopez, Amelie Hinrichsen	**Instrument-inspired controller** **Hardware:** Complex combination of sensors within the latex bellow - inertial accelerometer measures acceleration of the hand, further light sensors measure the degree of contraction and two microphones pick up airflow in and out of the valves (see Hinrichsen et al. 2014), plus Arduino microcontroller **Software:** SuperCollider, uses the *Influx Patch* for variable mapping configuration (Hinrichsen, Hardjowirogo, Hildebrand Marques Lopes & Bovermann, 2014)	**Causal** *Design and playing action inspired by the accordion, maps as expected onto output parameters, no movement = no sound*	2	Mode 2 Audio / Mode 1 Video, Mode 1 Audio / Mode 2 Video
S/A/S/A *(Fig. 3)*	Julius Fischötter	**Alternative controller** **Hardware:** 10 triangular fields of different materials, made conductive either through copper under-plates or through being mixed with iron casting powder and connected to laptop via a USB device, fields are pressed, struck, scratched etc.	**Causal** *Discrete/sustained actions mapped accordingly, no movement = no sound*	3	Mode 2 Audio / Mode 1 Video, Mode 3 Audio / Mode 2 Video, Mode 1 Audio / Mode 3 Video

Software: SuperCollider, with links to Ableton effect racks

Table 2

PCA Component loadings (n = 31)

	Rotated Component Loadings	
Item	**High positive Arousal**	**Low positive Arousal**
Exciting	**.92**	.12
Interesting	**.90**	.10
Attention Paid	**.89**	.17
Boring	**-.82**	-.12
Stimulating	**.82**	.25
Good	**.81**	.41
Calming	.01	**.95**
Pleasant	.44	**.82**
Eigenvalues	5.23	1.32
% of variance	65.4	16.5

Note. Component loadings > 0.5 are displayed in bold.

Figure Captions

Figure 1. The process of sound generation with a Digital Musical Instrument.

Figure 2. Venn diagram showing the distribution of words used in answers to survey Question 2, organized according to dimension of reception (Attention, Emotion, Liking).

Figure 3. The DMIs in the Causal Group. *Clockwise from top left: The Finger, PushPull and S/A/S/A*.

Figure 4. The DMIs in the Acausal Group. *Left: The Bass, Right: Jerry*.

Figure 5a-d. Interaction between Manipulation and Instrument Category for (a) Perceived Causality ratings, (b) Rating Component 1, (c) Rating Component 2, (d) Virtuosity ratings. *Brackets mark significant differences from t-tests, $p < .0125$ (Bonferroni correction).

Figure 6. A model of the DMI reception process.

Figure 7. Mean scores for the conditions *Causal Original, Causal Manipulated, Acausal Original* and *Acausal Manipulated* projected onto a 2-dimensional emotional valence and arousal space (cf. Russell, 1980). The original dimensions of High positive Arousal and Low positive Arousal were rotated by 90° to the right in order to have valence on the horizontal and arousal on the vertical dimension.

Figure 1.

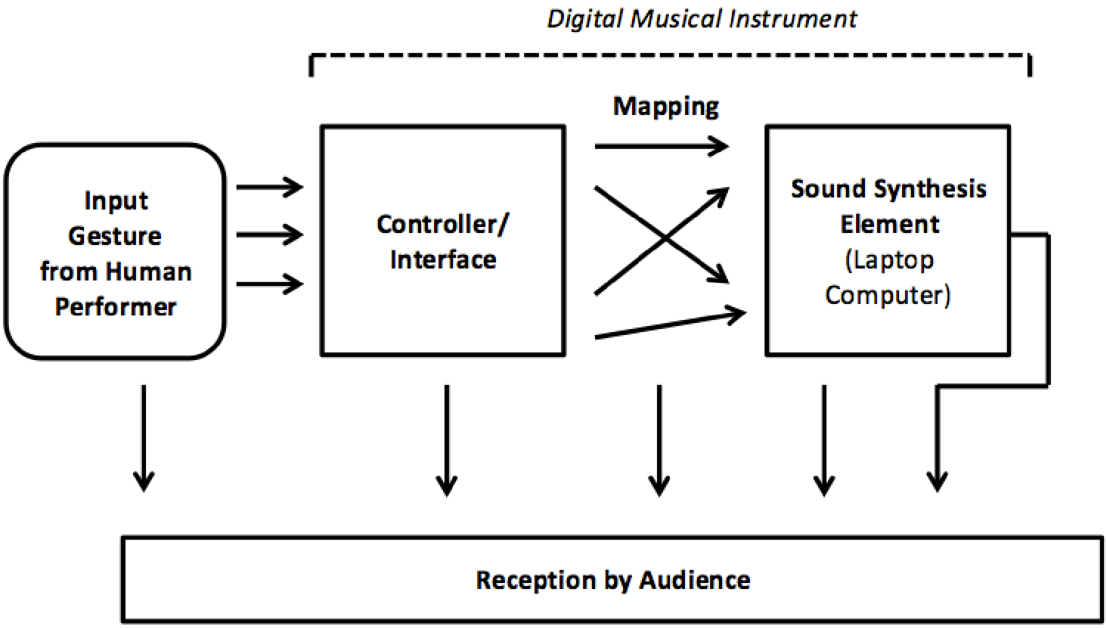

Digital Musical Instrument
Mapping
Input Gesture from Human Performer
Controller/ Interface
Sound Synthesis Element (Laptop Computer)
Reception by Audience

Figure 2.

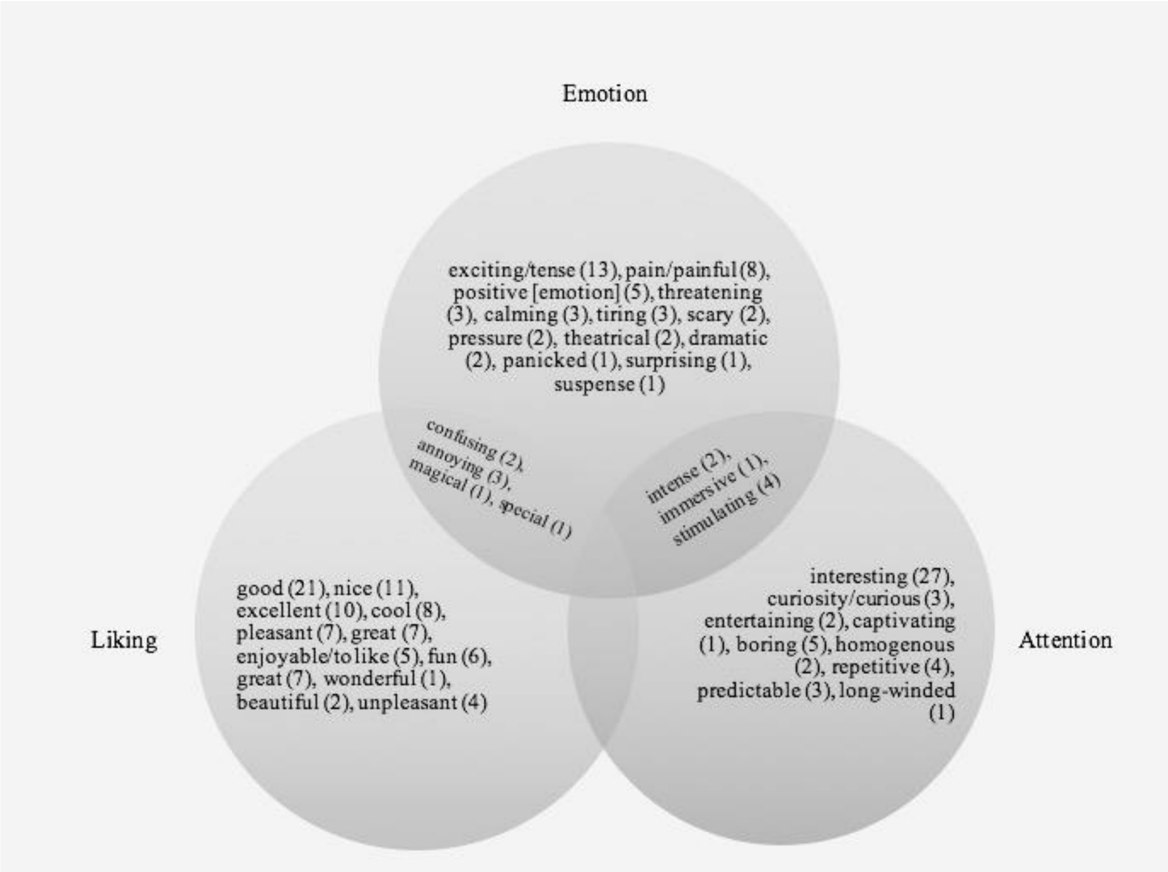

Emotion
exciting/tense (13), pain/painful (8), positive [emotion] (5), threatening (3), calming (3), tiring (3), scary (2), pressure (2), theatrical (2), dramatic (2), panicked (1), surprising (1), suspense (1)
confusing (2), annoying (3), magical (1), special (1)
intense (2), immersive (1), stimulating (4)
Liking
good (21), nice (11), excellent (10), cool (8), pleasant (7), great (7), enjoyable/to like (5), fun (6), great (7), wonderful (1), beautiful (2), unpleasant (4)
Attention
interesting (27), curiosity/curious (3), entertaining (2), captivating (1), boring (5), homogenous (2), repetitive (4), predictable (3), long-winded (1)

Figure 3.

Figure 4.

Figure 5.

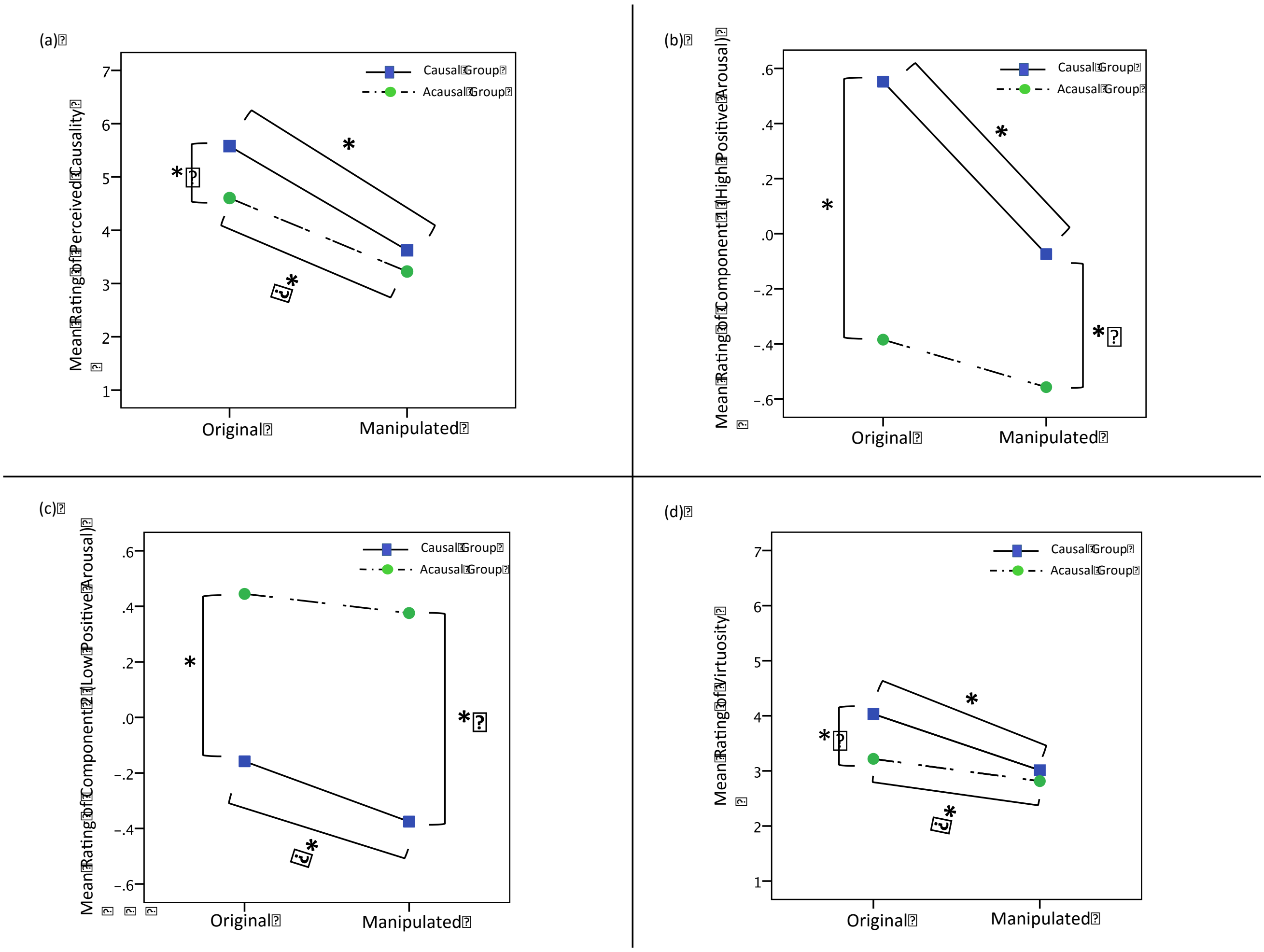

(a)
Mean Rating of Perceived Causality
Causal Group
Acausal Group
Original
Manipulated
(b)
Mean Rating of Component 1 (High Positive Arousal)
Causal Group
Acausal Group
Original
Manipulated
(c)
Mean Rating of Component 2 (Low Positive Arousal)
Causal Group
Acausal Group
Original
Manipulated
(d)
Mean Rating of Virtuosity
Causal Group
Acausal Group
Original
Manipulated

Figure 6.

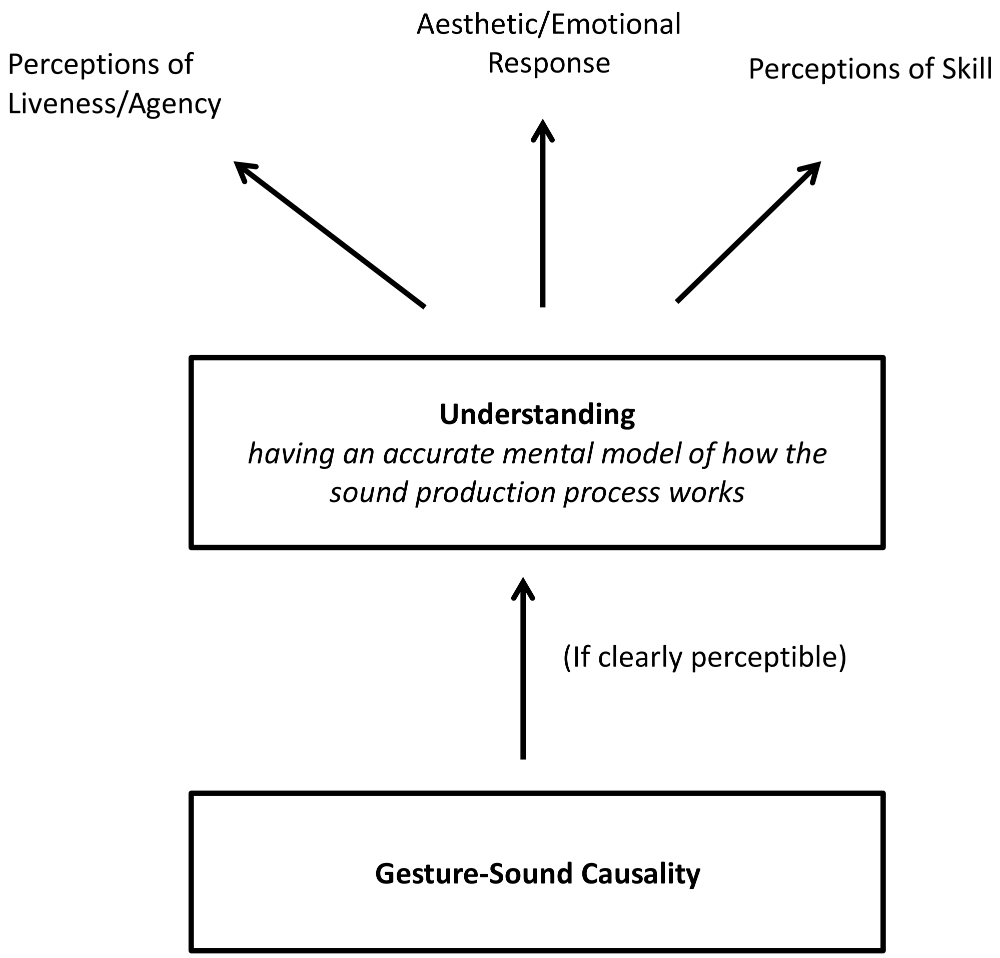
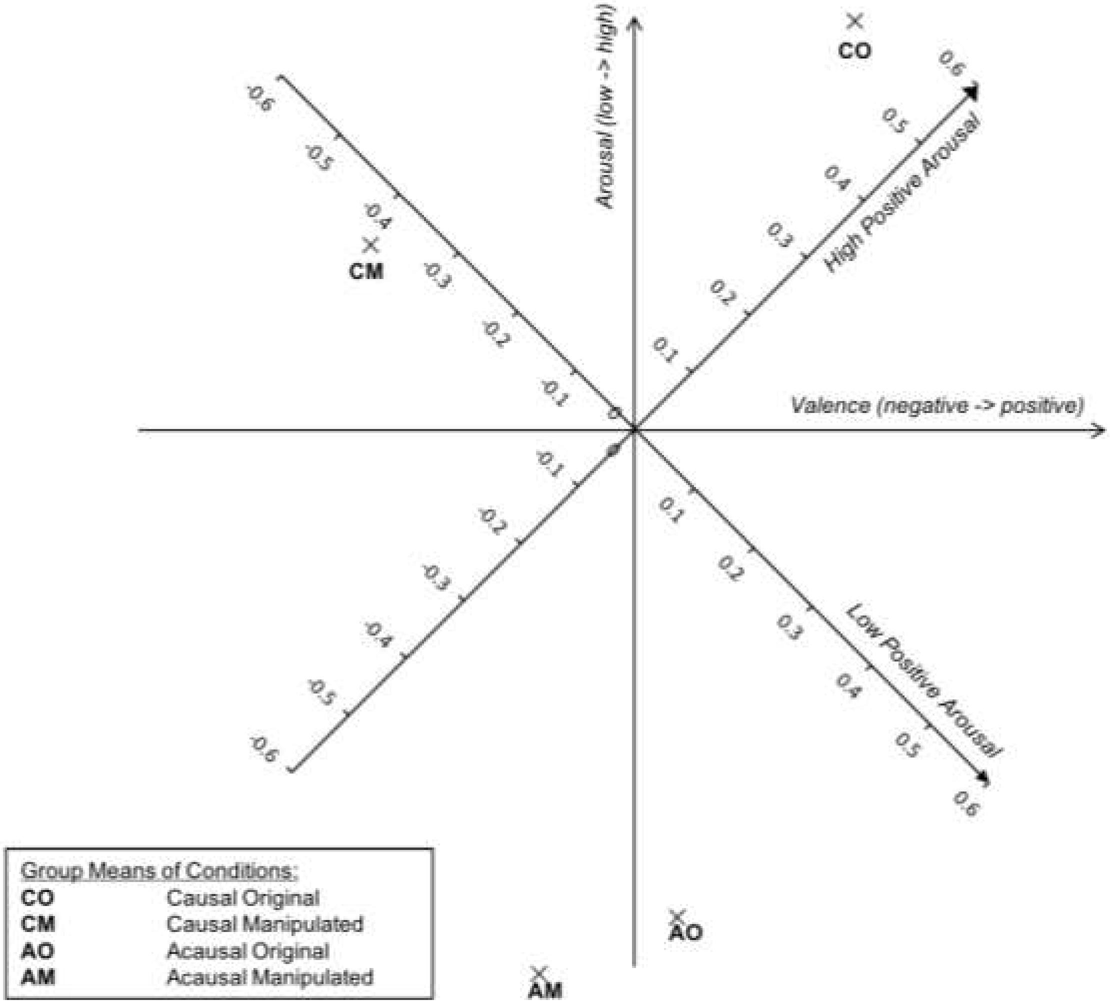

Perceptions of Liveness/Agency
Aesthetic/Emotional Response
Perceptions of Skill
Understanding
having an accurate mental model of how the sound production process works
(If clearly perceptible)
Gesture-Sound Causality

Figure 7.


Arousal (low -> high)
Valence (negative -> positive)
High Positive Arousal
Low Positive Arousal
CO
CM
AO
AM
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
Group Means of Conditions:
CO Causal Original
CM Causal Manipulated
AO Acausal Original
AM Acausal Manipulated

Appendix

Table 1

Results of Paired Sample t-Tests for Perceived Causality Ratings

	Condition Means	***SDs***	***df***	***t***	***r***
Causal Original vs. Acausal Original	5.58, 4.60	0.79, 1.39	30	5.67*	.72
Causal Manip. vs. Acausal Manip.	3.62, 3.23	1.11, 1.28	30	1.68	.29
Causal Original vs. Causal Manip.	5.58, 3.62	1.24, 1.14	30	5.89*	.73
Acausal Original vs. Acausal Manip.	4.60, 3.23	1.35, 1.19	30	4.49*	.63

Note. **p* < .0125 (Bonferroni corrected).

Table 2

Results of Paired Sample t-Tests for Component 1 (High Arousal, Positive Valence) Ratings

	Condition Means	***SDs***	***df***	***t***	***r***
Causal Original vs. Acausal Original	0.55, -0.39	0.55, 0.68	30	8.22^{*}	.83
Causal Manip. vs. Acausal Manip.	-0.07, -0.56	0.54, 0.65	30	4.93^{*}	.67
Causal Original vs. Causal Manip.	0.55, -0.07	0.55, 0.54	30	4.99^{*}	.67
Acausal Original vs. Acausal Manip.	-0.39, -0.56	0.68, 0.65	30	1.74	.30

Note. $^{*}p < .0125$ (Bonferroni corrected).

Table 3

Results of Paired Sample t-Tests for Component 2 (Low Arousal, Positive Valence) Ratings

	Condition Means	***SDs***	***df***	***t***	***r***
Causal Original vs. Acausal Original	-0.16, 0.45	0.49, 0.70	30	-5.33*	.69
Causal Manip. vs. Acausal Manip.	-0.38, 0.38	0.46, 0.73	30	-7.11*	.79
Causal Original vs. Causal Manip.	-0.16, -0.38	0.49, 0.46	30	4.67*	.64
Acausal Original vs. Acausal Manip.	0.45, 0.38	0.70, 0.73	30	0.86	.16

Note. **p* < .0125 (Bonferroni corrected).

Table 4

Results of Paired Sample t-Tests for Perceived Causality Ratings

	Condition Means	*SD*	*df*	*t*	*r*
Causal Original vs. Acausal Original	4.03, 3.22	1.24, 1.35	30	4.52*	.63
Causal Manip. vs. Acausal Manip.	3.01, 2.81	1.15, 1.19	30	1.46	.26
Causal Original vs. Causal Manip.	4.03, 3.01	1.24, 1.15	30	5.28*	.69
Acausal Original vs. Acausal Manip.	3.22, 2.81	1.35, 1.19	30	2.66*	.43

Note. *p < .0125 (Bonferroni corrected).