Causal Structures of Multidimensional QoE in Haptic-Audiovisual Communications: Bayesian Modeling

SHUJI TASAKA, Nagoya Industrial Science Research Institute

This article proposes a methodology for building and verifying plausible models that can express causation in multidimensional QoE for haptic-audiovisual interactive communications. For the modeling, we utilize subjective experimental data of five-point scores collected in a previous study where a pair of subjects carry out two kinds of interactive tasks (castanets hitting and object movement) in real space (not in virtual space). The multidimensional QoE is composed of 15 measures for the castanets hitting and 14 measures for the object movement. To reduce the dimension, we classify the QoE measures into three groups as indicators of three constructs (latent variables or factors): AVQ (AudioVisual Quality), HQ (Haptic Quality), and UXQ (User eXperience Quality). We then build two models: (1) a structural equation model in which AVQ and HQ correlated with each other give causal effects on UXQ, and (2) a confirmatory factor analysis model in which the three constructs are only correlated with each other. We refer to the former as 3C-SEM and the latter as 3C-CFA. We further introduce a CFA model with a single construct for which all QoE measures are its indicators (1C-CFA). We perform Bayesian analysis of the three models by means of Markov chain Monte Carlo simulation; in each model, the deviance information criterion is obtained for model comparison, and the posterior predictive p -value is calculated for model checking. As a result, we find that 3C-SEM is the most plausible and that HQ has a stronger causal effect on UXQ than AVQ. We also learn that the correlation between AVQ and UXQ is much higher than the direct causal effect and that the increase in the association as correlation is due to the causal effect of HQ on UXQ through the correlation of AVQ with HQ. Thus, it is suggested that improving haptic performance is more effective in enhancement of QoE than improving audiovisual performance.

CCS Concepts: • **Computing methodologies** → **Modeling methodologies**; • **Networks** → *Network performance modeling*; • **Human-centered computing** → *User models*;

Additional Key Words and Phrases: Quality of experience, QoE, causation, correlation, SEM, CFA, Bayesian modeling, latent variables, haptic-audiovisual interactive communications, MCMC, OpenBUGS

ACM Reference format:

Shuji Tasaka. 2020. Causal Structures of Multidimensional QoE in Haptic-Audiovisual Communications: Bayesian Modeling. *ACM Trans. Multimedia Comput. Commun. Appl.* 16, 1, Article 11 (February 2020), 23 pages.

<https://doi.org/10.1145/3375922>

This work was supported by JSPS KAKENHI (Grant-In-Aid for Scientific Research of Japan Society for the Promotion of Science) grant number 17K06454.

Author's address: S. Tasaka, Nagoya Industrial Science Research Institute, 1-13 Yotsuya-dori, Chikusa-ku, Nagoya, 464-0819, Japan; emails: tasaka@nisi.jp, s.tasaka@nitech.jp.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

1551-6857/2020/02-ART11 \$15.00

<https://doi.org/10.1145/3375922>

1 INTRODUCTION

Causality is a fundamental issue in science and technology [1–3]. This is also the case with multimedia communications quality of experience (QoE), which mainly consists of users' perceived quality of service (QoS) [4–6]. However, we can find no systematic approach to the causality problem of multimedia communications QoE in the literature; furthermore, related studies published so far are also limited.

This article presents a trial of developing a methodology for building and verifying plausible statistical models of *causation* (cause-effect relationship) in multimedia communications QoE. For the modeling, we consider *haptic-audiovisual* communications, which are regarded as one of the most promising multimedia services of the near future [7] and have traffic characteristics quite different from those of traditional audiovisual communications. As an example material of the modeling, we utilize subjective experimental data of five-point scores collected in a previous study on haptic-audiovisual *interactive* communications in real space (not in virtual space) [8].

Multimedia communications QoE is intrinsically multidimensional because it should reflect various aspects of the end-user experience of applications or services. Therefore, it usually has many measures, such as user perceived video quality, audio quality, and users' comfort and satisfaction. It is also apparent that the component measures can be correlated with each other (e.g., perceived video quality is correlated with satisfaction). Thus, multidimensional QoE modeling has to cope with *correlated multivariate data*, which are collected as scores (i.e., ordinal categorical data) of each component measure from subjects (users). This work assumes that QoE is represented by a *score variable*; we focus on the expectation of the score variable as the QoE measure.

It is well known in statistics that correlation is not causation [1, 2]. Correlation between two random variables A and B appears to be owing to either (1) causal effect of A on B , which is denoted as $A \rightarrow B$, (2) causal effect of B on A ($A \leftarrow B$), (3) reciprocal causal effect ($A \rightleftarrows B$), (4) a *confounder* (common cause) C of both A and B ($A \leftarrow C \rightarrow B$), or (5) others including combinations of (4) and one of the first three. We have to distinguish causation from correlation in multimedia communications QoE; the distinction enables us to utilize the causation to enhance QoE effectively so that human-centric services can be provided in an efficient way.

Issues of QoE in multimedia communications have a flavor of social and behavioral sciences since the human being is involved inside the target areas. Therefore, we can expect that some methodology in social and behavioral sciences can be useful for developing causation models of QoE in multimedia communications.

Structural equation modeling (SEM) is a powerful tool of this kind for modeling of correlated multivariate data and expressing causation between random variables [1–3, 9–13]. SEM originated in the field of genetics and was developed in non-engineering fields such as biometrics, econometrics, and psychometrics, and it had not attracted engineering researchers' attention.

When we employ SEM for QoE assessment, we regard the score variable as an *observed variable* (or an *indicator* or a *manifest variable*) in SEM terminology since it is directly observed by subjects.

To reduce the dimension for QoE modeling, SEM introduces *constructs*, each of which is a continuous latent variable indicating two or more observed variables with a similar property in a group as a factor of *factor analysis*. A construct cannot be observed directly; it is regarded as being measured indirectly through the corresponding observed variables. SEM formulates regression models between the constructs thus introduced, which is simple to analyze.

Note that application of traditional regression analysis to multivariate data of observed variables produces a large number of regression models because of many possible response variables, which makes it difficult to get the whole picture of interrelationship among the observed variables. In addition, simple regression models cannot clarify causal relationships between the response variable and explanatory variables [1, 2].

A SEM model, especially the LISREL type [13], consists of measurement equations and structural equations: the former is a set of equations relating observed variables (indicators) to constructs with loading coefficients, whereas the latter is regression models between endogenous and exogenous constructs.

The introduction of constructs yields not only the advantage of dimension reduction but also a disadvantage of analytic difficulty. A construct does not possess a natural metric, as it is a hypothetical variable: its location (origin), scale, and orientation are indeterminate [14], and therefore they must be resolved. This problem is referred to as identifiability [15, 16] or indeterminacies [14] in Bayesian analysis, and identification [9, 12] in the “frequentist” (classical) approach including maximum likelihood (ML) or least squares estimation.

The classical SEM, which assumes that the underlying distribution of the observed data is normal, works fine with reasonably large sample sizes but does not for small sizes [12]; however, the Bayesian SEM approach based on Markov chain Monte Carlo (MCMC) simulation can give more reliable results for smaller samples [10].

Thus, the Bayesian SEM approach can have advantages for multidimensional QoE modeling with small sample sizes, which are situations often encountered in subjective experiments. However, no study by this approach was found in the literature before 2016, when an earlier work of Tasaka [17] appeared, although Yamazaki et al. [18] reported a QoE study by classical SEM in 2012. An extended study of Tasaka [17] was published in 2019, which is a point of departure for the current study.

The remainder of this article is organized as follows. Section 2 gives a glance at related work and summarizes the novelty of this work. Section 3 briefs on the haptic-audiovisual interactive communication system used in this study and associated subjective experiments. Section 4 defines constructs and continuous latent variables underlying discrete scores, both of which we use to build the models. In Section 5, we build a SEM model with three constructs, which we call 3C-SEM, a confirmatory factor analysis (CFA) model with three constructs (3C-CFA) and a CFA model with a single construct (1C-CFA). Section 6 presents simulation results of the models thus built and compares them to find the most plausible one, in which causation is distinguished from correlation.

2 RELATED WORK AND NOVELTY OF THIS WORK

2.1 QoE Assessment

QoE assessment of haptic communications has been made extensively so far by many researchers (e.g., [7, 8, 17, 19, 20–25]). However, a Bayesian SEM approach can be found only in Tasaka [17, 19].

Many of the previous studies deal with virtual space or virtual reality (VR) (e.g., [20–22]). Hamam et al. [20] carried out subjective experiment on QoE assessment of a haptic handwriting learning system with a haptic device of *PHANToM Omni* [26] (hereafter, PHANToM for simplicity). To clarify the effects of three haptic factors (realism, usefulness, intuitiveness) and fatigue on overall QoE, Hamam et al. calculated the correlations and conducted statistical tests of significance, although mentioning that correlation does not necessarily mean causality. Another work by Hamam et al. [21] took a similar approach to QoE assessment of a tactile arm band device in addition to a force feedback system (PHANToM). Hamam et al. [22] also proposed a Fuzzy inference system (FIS) of QoE with five inputs (media sync, fatigue, haptic rendering, degree of immersion, and user intuitiveness) and examine the QoE output by calculating the correlation between the users’ responses and the FIS output and by doing hypothesis tests.

Regarding haptic communications in real space, Iwata et al. [23] investigated the effect of play-out buffering time on QoE in a haptic media, sound, and video transfer system with PHANToM, where the user of the master terminal beats the tambourine at the slave terminal; this is one-way

communication. Using the mean opinion score (MOS) as the QoE metric, they assessed the operability of PHANToM, sound output quality, video output quality, inter-stream sync quality and comprehensive quality; they also estimated the QoE measures by traditional regression analysis with explanatory variables of application-level QoS parameters. Tatematsu et al. [24] studied influences of network latency on QoE in the same system as that of Iwata et al. [23] in a similar way.

Causation in multidimensional QoE of haptic-audiovisual communications was first formulated by means of Bayesian SEM by Tasaka [17], where a pair of users perform an interactive task of object movement with PHANToM Omni. Since *playout buffering control* at the receiver (i.e., intra-stream synchronization control) improves the output temporal quality of individual media by absorbing network delay jitter, Tasaka evaluated two kinds of the control schemes: *MAB* (media adaptive buffering) and *BUF* (the conventional playout buffering control). The former sets the playout buffering time for the haptic media to a smaller value than that for the audio and video media, whereas the latter sets the same buffering time for the three media. A Bayesian SEM model was built with 14 observed variables of five-point scores (integers 1 through 5) approximated by continuous variables with discrete observations and hypothetical missing intermediate data (i.e., no model of categorical (discrete) data was developed). The model thus built classifies the 14 QoE measures into three groups, which are treated as indicators of three constructs (latent variables): *AVQ* (AudioVisual Quality), *HQ* (Haptic Quality), and *UXQ* (User eXperience Quality).

The Bayesian SEM model was analyzed by MCMC simulations with the SEM software *Amos* [27], and the QoE measures were examined quantitatively. As a result, MAB was shown to be superior to BUF from a QoE point of view, and it was demonstrated that the causal effect of HQ on UXQ is much larger than that of AVQ on UXQ.

The study in Tasaka [17] had several limitations and left problems to be solved. First, as already mentioned, the five-point score was treated approximately as a continuous variable. Second, we found a few parameters that did not satisfy the convergence statistics (CS) of MCMC simulation, which requires respecification of the model. Third, no issue of model checking was addressed.

Tasaka [19] resolved the limitations and problems mentioned earlier, taking over the framework of the SEM model proposed in an earlier work [17], especially the three constructs (AVQ, HQ, and UXQ); it adopts only MAB because of its superiority and deals with the task of object movement. For simplicity of modeling, Tasaka assumed that AVQ is uncorrelated with HQ [19]. He built Bayesian SEM models of categorical score data by introducing continuous latent variables underlying discrete scores in addition to the three constructs and formulated new measurement equations and structural equations. The models were analyzed by MCMC simulations with the software *OpenBUGS*, which is freely available [28, 29]. It has been shown that all parameters of the new models converge in the MCMC simulations. Tasaka then performed model checking with a *deviance information criterion* (DIC) and *posterior predictive p-value* (PPp-value) of a discrepancy function. He also provided systematic treatment of the identifiability issue of the latent variables in QoE by resorting to two methods for scaling constructs: *unit loading* (UL) and *unit variance* (UV).

2.2 Identifiability in Latent Variables

A key to modeling with latent variables is the identifiability of location, scale, and orientation. The problem is resolved by fixing various parameters [9–12, 14–16, 27].

The location indeterminacy is resolved by fixing the means of latent variables to a certain value, usually 0.

Typical scaling methods are (1) UL and (2) UV. UL is to fix a loading on each construct, usually to 1, whereas UV fixes the variance of the construct to 1.

The orientation issue implies that a construct may be rotated 180° or “flipped” [14] (i.e., labeling shifts (label switching) [16]), since the product of a construct Ψ and a loading λ ($\Psi \times \lambda$) is indistinguishable from $(-\Psi) \times (-\lambda)$. This issue can be addressed by fixing loadings in some ways.

The identification issue in the frequentist approach has been extensively studied, and various identification rules have been established [9, 12]. Bayesian approaches, however, have not yet been thoroughly investigated, but they are more flexible; models that are not identified in a frequentist approach can often be specified by Bayesian ones [14]. The identifiability can be checked by mathematical methods or empirical ones. This work utilizes an empirical method by examining MCMC simulation results.

As will be seen in Section 5, many mechanisms used to resolve the indeterminacies may be seen as *prior information* [14]. We can get the priors from expert knowledge of the target system to be modeled, which network engineers should have.

2.3 Novelty of This Work

The study in this article started with the results obtained in Tasaka [19]. The novelty of this work can be summarized as follows.

For the task of *castanets hitting* in addition to the object movement, we build Bayesian SEM models with the three constructs in which AVQ and HQ are correlated and give causal effects on UXQ; this type model is referred to as 3C-SEM in this article. To show the plausibility of the existence of the causation, we also build a CFA model in which the three constructs are only correlated with each other (3C-CFA model) and a CFA model with a single construct for which all of the QoE measures are its indicators (1C-CFA model) for each of the two tasks. We further compare the 3C-SEM, 3C-CFA, and 1C-CFA models in each task to show that 3C-SEM is the most plausible in the sense that it has the minimum value of DIC, while keeping the PPp-values within a satisfactory range of goodness of fit. We also quantify correlation between the constructs and discuss the difference in magnitude between the correlation coefficient and the causal effect. This is the first study on quantitative discrimination between causation and correlation in QoE of haptic-audiovisual interactive communications.

OpenBUGS codes of all of the models built in this work are provided in the supplementary material accompanying this article.

3 HAPTIC-AUDIOVISUAL INTERACTIVE COMMUNICATIONS

For QoE modeling, this work utilizes a subset of the experimental data collected in Isomura et al. [8].

A pair of *subjects* each using a haptic device (PHANToM) performs two kinds of interactive tasks, *castanets hitting* and *object movement*, for which one subject physically manipulates the other subject’s PHANToM remotely.

This section briefly describes the experimental system and subjective experimental data we will use for the modeling.

3.1 Experimental System

Figure 1 illustrates the experimental system configuration, the workspace for the task of castanets hitting (right top), and that for the task of object movement (right bottom).

Each terminal is equipped with a haptic device (PHANToM), a Web camera, an LCD monitor, and a headset, and has an identical workspace, where a subject carries out the interactive task. The two terminals are located in different rooms. Each terminal transmits/receives haptic media, audio, and video to/from the other terminal as three separate UDP streams. The transmission links are full duplex Ethernet channels of 100 Mb/s or 10 Mb/s.

distributed intervals (i.e., Poisson traffic). The average bit rate of the load traffic is set to 6.0 Mb/s as an appropriate amount of interference traffic for study purposes [8, 25].

3.2 Procedures for Doing the Tasks

In a task, a subject plays one of two roles exclusively: the *instructor* and the *manipulator*. The task is done in the following three steps repeatedly during an experiment run, which is defined as 30 seconds:

- Step 1:* The instructor gives the manipulator an audio instruction (the number of hitting in the castanets hitting task or the target object and its destination in the object movement task).
- Step 2:* The manipulator takes the action specified by the instruction.
- Step 3:* Once the manipulator has completed the action, the two subjects alternate the roles (manipulator $\rightarrow$ instructor, and instructor $\rightarrow$ manipulator) and return to Step 1.

During the action, the instructor's subject holds his or her own stylus in a hand and surrenders himself or herself to the manipulator's movement not so as to impede it, just as he or she looks like a marionette remotely operated by the manipulator. Note that the instructor's stylus is geared to the manipulator's stylus over the network because of a force feedback loop of PHANToM.

(1) *Castanets hitting.* The procedure for the task of castanets hitting is as follows.

As illustrated at the top of Figure 1(b), a pair of castanets is set just under the stylus of PHANToM in each workspace. The castanets, the stylus, and the subject's forearm are put in the visual field of the Web camera.

First, the instructor randomly selects the number of hitting from 1 to 3 and announces the number to the manipulator using the microphone as an audio instruction.

According to the instruction, the manipulator swings his or her own stylus down to the castanets from a height of about 15 cm at regular intervals specified in advance so that the stylus at the instructor's side, as well as the manipulator's side, can hit the castanets. The sound of the hitting is captured by the microphone of the headset and transmitted to the other terminal. Therefore, each subject can confirm the hitting of not only his or her own but also the other's hitting.

Once the manipulator has hit the castanets the number of times specified by the instruction, the two subjects alternate the roles. The preceding procedure is repeated during an experiment run.

(2) *Object movement.* The instructor randomly selects an object put in the center circle and its destination out of the three kinds (a circle, a triangle, or a square) on his or her own side (see the bottom of Figure 1(b)) and announces the selections using the microphone (e.g., move the triangle object to the circle destination). The manipulator moves the specified object to the specified destination on the instructor's workspace with the instructor's PHANToM stylus by manipulating his or her own stylus while watching the instructor's workspace displayed on his or her LCD monitor.

Note that the two tasks have contrastive traits. *Fast response* is essential for the castanets hitting, whereas this is not necessarily the case with the object movement. However, the object movement requires precise operation of PHANToM with video monitoring of the workspace at the other side; this requirement is not so strict for the castanets hitting. The time taken for hitting the castanets once is approximately 0.7 to 1 s, and the stylus moving speed is about 300 to 500 mm/s. This implies a large number of the role alternation during an experimental run of 30 seconds. When the end-to-end delay between the two terminals becomes long, the subjects may recognize asynchrony between their own castanets' sound and the other one. Regarding the object movement, the time taken for moving an object is approximately 10 seconds on average,

Table 1. Subjective QoE Measures and Score Criteria for the Castanets Hitting and Object Movement

Subjective QoE Measure (observed variable)	Score Criteria (5-point scale) 1 (leftmost), 2, 3 (neutral), 4, 5 (rightmost)	Task	
		Castanets Hitting	Object Movement
		Score (latent variable)	Score (latent variable)
Video Temporal quality	Video is jerky $\leftrightarrow$ smooth	$z_1(v_t)$	$z_1(v_t)$
Video Spatial quality	Video is corrupt $\leftrightarrow$ clear	$z_2(v_s)$	$z_2(v_s)$
Video Usefulness	Image is hard to grasp $\leftrightarrow$ easy	$z_3(v_u)$	$z_3(v_u)$
Inter-stream sync between video and haptic	Out of sync $\leftrightarrow$ In sync	$z_4(v_h)$	$z_4(v_h)$
Audio Naturalness	Audio is artificial $\leftrightarrow$ natural	$z_5(a_n)$	$z_5(a_n)$
Castanet Sync	Out of sync $\leftrightarrow$ In sync	$z_6(c_s)$	Not applicable
Haptic Smoothness	Movement is awkward $\leftrightarrow$ smooth	$z_7(h_m)$	$z_6(h_m)$
Haptic Operability	Manipulation is heavy $\leftrightarrow$ light	$z_8(h_o)$	$z_7(h_o)$
Haptic Stability	Manipulation is unstable $\leftrightarrow$ stable	$z_9(h_s)$	$z_8(h_s)$
Response	Response is slow $\leftrightarrow$ quick	$z_{10}(r)$	$z_9(r)$
Communicability	Hard to communicate $\leftrightarrow$ Easy	$z_{11}(c_m)$	$z_{10}(c_m)$
System comfort	Uncomfortable $\leftrightarrow$ Comfortable	$z_{12}(c_f)$	$z_{11}(c_f)$
Work difficulty	Difficult to work $\leftrightarrow$ Easy	$z_{13}(w_d)$	$z_{12}(w_d)$
Overall satisfaction	Unsatisfied $\leftrightarrow$ Satisfied	$z_{14}(s_o)$	$z_{13}(s_o)$

although it depends on individual subjects, and the stylus moving speed is about 30 to 60 mm/s. Thus, the number of the role alternation during an experimental run is small.

3.3 Subjective Experimental Data

In this section, we describe the QoE measures and score criteria, focusing on the task of castanets hitting. Excluding a QoE measure (*Castanet Sync*) from those of the castanets hitting leads to the case of the object movement.

For each of the six stimuli in the castanets hitting, we have the experimental data on five-point scores of 14 subjective QoE measures and an objective QoE measure, which is the number n_c of instructions given by the instructor to the manipulator during an experimental run (30 seconds) and is considered a sort of task efficiency.

The left three columns of Table 1 list the 14 subjective QoE measures for the castanets hitting; the third column in a row defines the notation z_j ($j = 1, \dots, 14$) of the score variable corresponding to the QoE measure in the row.

In this work, we are planning to build a Bayesian model explicitly formulating *ordinal categorical scores* $z_1, \dots, z_{14}$, which are observed variables. For that purpose, we resort to a latent variable method proposed in Tasaka [30]. For each observed variable, we introduce an *underlying continuous latent variable*, whose notation is defined inside the parentheses in the third column of Table 1 so as to be the same as that in other works [17, 19].

The subjects participating in this experiment were 42 Japanese people (16 male and 26 female) in their teens to their 50s. According to the score criteria specified in the top row in the second column of Table 1, each subject evaluated the six stimuli. The stimuli were presented in random order to the subject pair (additionally, the order of the task to be done was randomized on the subject-pair basis).

For convenience, the rightmost column of Table 1 shows the score variables and the corresponding latent variables in the object movement. Note that in the bottom eight rows, the subscript j

Table 2. Latent Variables and Corresponding Indicators for Subject i ($i = 1, \dots, N$; $N = 42$): Castanets Hitting

Latent Variable	Corresponding Indicators for Castanets Hitting
f_{i1} (AVQ)	$z_{i1}^*, z_{i2}^*, z_{i3}^*, z_{i4}^*, z_{i5}^*, z_{i6}^*$
f_{i2} (HQ)	$z_{i7}^*, z_{i8}^*, z_{i9}^*$
f_{i3} (UXQ)	$z_{i,10}^*, z_{i,11}^*, z_{i,12}^*, z_{i,13}^*, z_{i,14}^*, z_{i,15}^*$
$z_{ij}^* \quad j = 1, \dots, 15$	$z_{ij} \quad j = 1, \dots, 15$

of the score variable z for the object movement is smaller by 1 than that for the castanets hitting, since Castanet Sync is not applicable to the object movement.

4 CONSTRUCTS AND UNDERLYING LATENT VARIABLES

4.1 Definition of Constructs

To build Bayesian models for ordinal categorical data, we first group all of the observed variables into a certain number of constructs (latent variables or factors). As mentioned in Section 2.3, this work considers two cases: three constructs and a single construct. In this section, we discuss only the task of castanets hitting; the case of the object movement parallels this discussion.

In the case of three constructs, we classify the observed variables into three groups as shown in Table 1; we introduce two *exogenous* constructs, AVQ and HQ, which are the first two groups, and an *endogenous* construct, UXQ, which is the last group. We denote these *latent variables* as f_1, f_2 and f_3 , respectively.

Modeling of cause-effect relationships needs some assumptions provided by humans, resting on expert judgment [1]. According to this principle, we assume that both AVQ and HQ affect UXQ (i.e., $AVQ \rightarrow UXQ$ and $HQ \rightarrow UXQ$) and that AVQ and HQ are correlated. This is because it is reasonable for the former assumption to consider that components affect the whole, and for the latter assumption, *cross modality* between audio-video and haptic can be expected. We refer to this model as 3C-SEM.

For comparison purposes, we also make two other assumptions of the constructs as mentioned earlier: the three constructs only correlated with each other (3C-CFA) and a single construct (1C-CFA).

For modeling convenience, we classify n_c , which is the number of instructions given by the instructor, into UXQ. If the count of n_c exceeds 5, then it is truncated; n_c is set to 5. We will examine the accuracy of the approximation later. This modified version of n_c is redefined as z_{15} .

Regarding the 15 underlying continuous latent variables for the observed variables z_j 's, we define a set of aliases:

$$\begin{aligned} & (z_1^*, z_2^*, z_3^*, z_4^*, z_5^*, z_6^*, z_7^*, z_8^*, z_9^*, z_{10}^*, z_{11}^*, z_{12}^*, z_{13}^*, z_{14}^*, z_{15}^*) \\ & \triangleq (v_t, v_s, v_u, v_h, a_n, c_s, h_m, h_o, h_s, r, c_m, c_f, w_d, s_o, n_c). \end{aligned}$$

To proceed to the next step of Bayesian modeling, we should now pay attention to the fact that the latent variables f_n ($n = 1, 2, 3$), z_j^* 's, and the observed variables z_j 's depend on *individual subjects*. We therefore add a subscript i , which represents subject i ($i = 1, \dots, N$; $N = 42$ in this study), to the latent and observed variables. Consequently, for subject i ($i = 1, \dots, N$), we employ notations $f_{i1}, f_{i2}, f_{i3}, z_{ij}^*, z_{ij}$ ($j = 1, \dots, J$; $J = 15$ for the castanets hitting¹).

Table 2 summarizes relations between the latent variables introduced so far and the corresponding indicators (observed variables). It should be noted that the observed variables $z_{i1}, \dots, z_{i,15}$ are

¹ $J = 14$ for the object movement.

generated by the underlying continuous latent variables $z_{i1}^*, \dots, z_{i,15}^*$, which serve as the observed (but actually *unobserved*) variables of the constructs f_{i1}, f_{i2}, f_{i3} , which explains the correlations between the z_{ij}^* variables. Thus, we are faced with a situation in which

$$\text{unobserved}(f_{in}) \rightarrow \text{unobserved}(z_{ij}^*) \rightarrow \text{observed}(z_{ij}).$$

This is a hard problem for statistical inference and is referred to as doubly missing data [15]. The preceding situation suggests that the key to the overall identifiability is resolving indeterminacies of z_{ij}^* first.

4.2 Underlying Continuous Latent Variables

We resolve indeterminacies of z_{ij}^* by utilizing the method proposed in Tasaka [30]. The observed variable z_{ij} , which represents categorical data $1, 2, \dots, K$ ($K = 5$), can be related to the underlying continuous latent variables z_{ij}^* with $K - 1$ *cut points* (category boundaries) κ_{jk} as

$$z_{ij} = k, \quad \text{if } \kappa_{j,k-1} < z_{ij}^* \leq \kappa_{jk}; \quad k = 1, \dots, K, \quad (1)$$

where $-\infty = \kappa_{j0} < \kappa_{j1} \leq \dots \leq \kappa_{j,K-1} < \kappa_{jK} = \infty$. Then, it is shown that

$$p_{ij}(k) = F(\kappa_{jk} - \mu_{ij}); \quad k = 1, \dots, K - 1, \quad (2)$$

where $p_{ij}(k)$ denotes the *cumulative probability* of z_{ij} being k or lower, and $F(\varepsilon_{ij})$ is the *probability distribution function* of ε_{ij} , which is the *deviation* of z_{ij}^* ; $\varepsilon_{ij} \triangleq z_{ij}^* - \mu_{ij}$, and $\mu_{ij} = E[z_{ij}^*]$ for given i and j .

Since the logistic function achieves fast and stable convergence in MCMC simulation of Bayesian models, we set $F(\varepsilon_{ij}) = \text{logistic}(\varepsilon_{ij}) = 1/[1 + \exp\{-(\pi/\sqrt{3})\varepsilon_{ij}\}]$; note that the variance of ε_{ij} is 1 because of the coefficient of $\pi/\sqrt{3}$ [28]. We thus have

$$p_{ij}(k) = 1/[1 + \exp\{-(\pi/\sqrt{3})(\kappa_{jk} - \mu_{ij})\}]; \quad k = 1, \dots, K - 1, \quad p_{ij}(K) = 1. \quad (3)$$

We let μ_{ij} load on either f_{i1} , f_{i2} , or f_{i3} according to the classification given in Table 2 (e.g., μ_{i1} loads on f_{i1} , and μ_{i7} loads on f_{i2}). This can resolve the orientation indeterminacy of the constructs f_{i1} , f_{i2} , and f_{i3} [14]. The details will be specified as a *non-overlapping structure* in the next section.

4.3 Resolving Scale and Orientation Indeterminacies of Constructs

As already mentioned, this problem typically can be resolved with UL or UV.

The UL method fixes the mean of the construct to 0 and the loading for the first observed variable to 1, while freeing the restriction on the loading coefficients of the other observed variables for the construct—that is, it assigns the construct the same unit as one of the observed variables, and the variances of the constructs are free parameters. In particular, the loading coefficient of +1 for the first observed variable of each construct can resolve the orientation indeterminacy and as the scaling one [16].

The UV method fixes the variances of the *exogenous* construct to 1 along with zero mean, which sets the scale to 1, and all of the loading coefficients are free parameters. However, positive values of the loading coefficients are needed for resolution of the orientation indeterminacy. Note that the variance of an *endogenous* construct is not a free parameter.

In Tasaka [19], we also find that the UL method is somewhat better than the UV method in a Bayesian context in terms of the DIC and PPp-value. Therefore, we apply the UL method to both 3C-SEM and 3C-CFA models while employing the UV method for 1C-CFA for simplicity of modeling. In the next section, we build the three Bayesian models for the task of castanets hitting. The Bayesian models for the object movement can be obtained by excluding the variable of Castanet Sync.

Fig. 2. Illustrative diagrams of latent variable models for the castanets hitting. (a) The structural model (construct model) + measurement model for 3C-SEM. (b) The construct models for 3C-CFA (top) and 1C-CFA (bottom). For simplicity, subject index i is omitted in all variables.

5 BAYESIAN CATEGORICAL MODELS FOR THE CASTANETS HITTING

First of all, we define the *QoE* as the expectation of the scores and let $\text{qoe}[j]$ denote the *QoE* for the j -th observed variable ($j = 1, \dots, 15$). Defining $\pi_{ij}(k) \triangleq \Pr(z_{ij} = k | i, j)$, we have $\pi_{ij}(1) = p_{ij}(1)$, $\pi_{ij}(k) = p_{ij}(k) - p_{ij}(k-1); k = 2, \dots, K$, and can express $z_{ij} \sim \text{Categorical}(\pi_{ij}(1), \dots, \pi_{ij}(K))$. Then, $\text{qoe}[j]$ is given by

$$\text{qoe}[j] = \sum_{i=1}^N \sum_{k=1}^K k \cdot \pi_{ij}(k) / N. \quad (4)$$

In the following, we build 3C-SEM, 3C-CFA, and 1C-CFA models in turn.

Figure 2 depicts illustrative diagrams of the latent variable models. The left panel (Figure 2(a)) shows the whole model of 3C-SEM consisting of the measurement model and structural model, which we also call the *construct model* in this article. The right panel (Figure 2(b)) represents only construct models for 3C-CFA (top) and 1C-CFA (bottom). It should be noted that these models are non-linear owing to the categorization of scores.

5.1 3C-SEM Model Based on the UL Method

We first set the means of the exogenous constructs (AVQ and HQ) to 0: $E[f_{i1}] = 0$ and $E[f_{i2}] = 0$.

Next, let λ_{jn} ($j = 1, \dots, 15; n = 1, 2, 3$) denote the loading coefficient of μ_{ij} on the construct f_{in} ; note that the loading coefficients are assumed not to depend on i . In addition, Λ is defined as a 15×3 matrix of the loading coefficients of an individual subject. We then assume that Λ has a non-overlapping structure of the coefficients [10] because of the orientation indeterminacy constraint [14] and the causality assumption ($\text{AVQ} \rightarrow \text{UXQ} \leftarrow \text{HQ}$). This structure can be regarded as representation of informative priors that come from expert knowledge of the haptic-audiovisual communication system.

Since the first observed variable is z_{i1} for the construct f_{i1} , z_{i7} for the construct f_{i2} , and $z_{i,10}$ for the construct f_{i3} , we set $\lambda_{11} = \lambda_{72} = \lambda_{10,3} = 1$. The non-overlapping structure also specifies $\lambda_{j1} = 0$ ($j = 7, \dots, 15$), $\lambda_{j2} = 0$ ($j = 1, \dots, 6, 10, \dots, 15$), and $\lambda_{j3} = 0$ ($j = 1, \dots, 9$). Setting of $\lambda_{71} = 0$, for instance, implies that z_{i7} (haptic smoothness) does not depend on f_{i1} (AVQ). Note that the 3C-SEM model in Figure 2(a) exhibits the non-overlapping structure.

We have also evaluated a case where $\lambda_{42} \neq 0$, which means that z_{i4} (v_h) depends on f_{i2} in addition to f_{i1} via $\lambda_{41} \neq 0$. We have then found no substantial difference between the two cases.

5.1.1 Measurement Equations and Structural Equation. On the basis of the preceding description, the measurement equations are expressed as follows:

$$p_{ij}(k) = 1 / \left[1 + \exp \left\{ - \left(\pi / \sqrt{3} \right) \left(\kappa_{jk} - \mu_{ij} \right) \right\} \right]; \quad k = 1, \dots, 4 \quad (5)$$

$$\mu_{i1} = f_{i1} \quad (6)$$

$$\mu_{ij} = \lambda_{j1} \cdot f_{i1}; \quad j = 2, \dots, 6 \quad (7)$$

$$\mu_{i7} = f_{i2} \quad (8)$$

$$\mu_{ij} = \lambda_{j2} \cdot f_{i2}; \quad j = 8, 9 \quad (9)$$

$$\mu_{i,10} = f_{i3} \quad (10)$$

$$\mu_{ij} = \lambda_{j3} \cdot f_{i3}; \quad j = 11, \dots, 15. \quad (11)$$

The structural equation given f_{i1} and f_{i2} might be written as

$$f_{i3} = \gamma_1 \cdot f_{i1} + \gamma_2 \cdot f_{i2} + \delta_i, \quad (12)$$

where the regression coefficients γ_1 and γ_2 represent causal effects $\text{AVQ} \rightarrow \text{UXQ}$ and $\text{HQ} \rightarrow \text{UXQ}$, respectively, and δ_i is the residual error following a normal distribution with zero mean and a variance of σ_δ^2 : $\delta_i \sim N(0, \sigma_\delta^2)$. Thus, the conditional distribution of f_{i3} given f_{i1} and f_{i2} is normal:

$$f_{i3} | f_{i1}, f_{i2} \sim N(v_i, \sigma_3^2); \quad v_i = \gamma_1 \cdot f_{i1} + \gamma_2 \cdot f_{i2}. \quad (13)$$

The joint distribution of f_{i1} and f_{i2} is assumed to be a bivariate normal distribution with zero marginal means; marginal variances σ_1^2 and σ_2^2 , respectively; and correlation $\rho = \text{Cov}[f_{i1}, f_{i2}] / (\sigma_1 \sigma_2)$. Note that the conditional distribution $f_{i2} | f_{i1}$ is also normal with mean $(\sigma_2 / \sigma_1) \rho f_{i1}$ and variance $(1 - \rho^2) \sigma_2^2$; this is useful for writing OpenBUGS codes [28, 29], where the *precision* $\tau_n \triangleq 1 / \sigma_n^2$ ($n = 1, 2, 3$) is employed instead of the variance σ_n^2 .

We next give *priors*. Priors of the cut points are set as in Tasaka [30]: $\kappa_{j1} \sim N(0, 1)$, $\kappa_{jk} = \kappa_{j,k-1} + \Delta_{jk}$, and $\Delta_{jk} \sim \text{Exponential}(1)$ for $k = 2, \dots, K - 1$, where the mean is 1. Note that this setting ensures the cut points satisfying the ordinal relations of the scores 1, 2, $\dots$, 5. Regarding loadings, we employ *relatively informative priors* [16]: $\lambda_{j1} \sim N(1, 1)$ ($j = 2, \dots, 6$), $\lambda_{j2} \sim N(1, 1)$ ($j = 8, 9$), and $\lambda_{j3} \sim N(1, 1)$ ($j = 11, \dots, 15$). This is because there is no particular reason to expect widely differing loading coefficients between the indicators [16]. The regression coefficients are supposed to be equally weighted: $\gamma_1 \sim N(0.5, 1/\tau_3)$ and $\gamma_2 \sim N(0.5, 1/\tau_3)$.

As for the precisions τ_1, τ_2 , and τ_3 , we adopt *hierarchical priors*: $\tau_n = \sigma_n^{-2}$ ($n = 1, 2, 3$), $\sigma_n \sim \text{Uniform}(0, 10)$, which represents a uniform distribution in $[0, 10]$ and is a *hyperprior* of *Gelman-style vague uniform distribution* of standard deviation.

5.1.2 Standardized Coefficient. To compare the causal effects represented by γ_1 and γ_2 under a reasonable condition, we first consider the meanings of these regression coefficients. The γ_1 coefficient is interpreted as the expected change in f_{i3} for a one-unit change in f_{i1} , keeping f_{i2} constant, and the γ_2 is understood similarly. This leads to a difficulty of the comparison between γ_1 and γ_2 , since f_{i1} and f_{i2} have different units of scales (i.e., different variances).

The *standardized coefficient*, which is the regression coefficient of a *standardized variable* (obtained by subtracting the mean from the original variable and dividing it by the standard deviation), is useful for assessing relative effects of different explanatory variables. The standardized coefficient shows the mean response in standard deviation units of the response variable for a one standard deviation change in an explanatory variable, holding constant the other variables in a model [9].

According to Bollen [9], the standardized coefficients of the regression coefficients, denoted as $\gamma_{s,n}$ ($n = 1, 2$), are formulated as $\gamma_{s,n} = \gamma_n \sigma_n / \sqrt{\text{Var}[f_{i3}]}$ ($n = 1, 2$). Because $\text{Var}[f_{i3}|f_{i1}, f_{i2}] = \sigma_3^2$, the unconditional variance of f_{i3} is calculated as $\text{Var}[f_{i3}] = E[(\gamma_1 f_{i1} + \gamma_2 f_{i2} + \delta_i)^2] = \gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 + 2\gamma_1 \gamma_2 \sigma_1 \sigma_2 \rho + \sigma_3^2$, where we assume that $E[f_{i1} \delta_i] = 0$ and $E[f_{i2} \delta_i] = 0$. We thus obtain

$$\gamma_{s,n} = \gamma_n \sigma_n / \sqrt{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 + 2\gamma_1 \gamma_2 \sigma_1 \sigma_2 \rho + \sigma_3^2}, \quad n = 1, 2. \quad (14)$$

Similarly, the standardized loading coefficient can be calculated as

$$\lambda_{s,j1} = \lambda_{j1} \sigma_1; \quad j = 1, \dots, 6 \quad (15)$$

$$\lambda_{s,j2} = \lambda_{j2} \sigma_2; \quad j = 7, 8, 9 \quad (16)$$

$$\lambda_{s,j3} = \lambda_{j3} \sqrt{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 + 2\gamma_1 \gamma_2 \sigma_1 \sigma_2 \rho + \sigma_3^2}; \quad j = 10, \dots, 15. \quad (17)$$

Using the standardized coefficients thus obtained, we can now calculate the correlation coefficients between the constructs $\rho_{13}^{(SEM)}$ for $AVQ \leftrightarrow UXQ$ and $\rho_{23}^{(SEM)}$ for $HQ \leftrightarrow UXQ$ as follows:

$$\rho_{13}^{(SEM)} = \gamma_{s,1} + \rho \cdot \gamma_{s,2} \quad (18)$$

$$\rho_{23}^{(SEM)} = \gamma_{s,2} + \rho \cdot \gamma_{s,1}. \quad (19)$$

Of course, we have $\rho_{12}^{(SEM)} = \rho$ for $AVQ \leftrightarrow HQ$. Although we can derive Equations (18) and (19) by algebraic manipulation, *Wright's tracing rules* in the *path analysis* [12, 31] of the structural model easily provide the same results.

It should be noted that the first term on the right-hand side of each of Equations (18) and (19) represents the *direct causal effect* on UXQ , whereas the second term implies *correlation due to the indirect path* between the two variables via correlation between AVQ and HQ .

5.2 3C-CFA Model Based on the UL Method

The whole model of 3C-CFA can be obtained by replacing the construct model of Figure 2(a) with the top panel in Figure 2(b), which produces a CFA model with three factors. We adopt the UL method and the loading matrix Λ with the non-overlapping structure. Therefore, the measurement equations are the same as those for 3C-SEM: Equations (5) through (11). The corresponding priors are also the same as those for 3C-SEM.

For the construct model, the joint distribution of f_{i1} , f_{i2} and f_{i3} is assumed to be *multivariate normal* with zero means: $MVN(0, V)$, where V is a 3×3 symmetric matrix as the variance-covariance matrix whose prior distribution must be specified in Bayesian analysis. In OpenBUGS, the prior of the *precision matrix* $P \triangleq V^{-1}$ is used instead; both P and V must always be *positive definite*.

It is often the case that the *conjugate prior* distribution is selected in Bayesian analysis because of analytical simplicity [10, 16, 28]. The conjugate prior of the multivariate normal distribution is *Wishart* [10, 28]. Thus, P is assumed to follow three-dimensional Wishart distribution:

$$(f_{i1}, f_{i2}, f_{i3}) \sim MVN(0, V); \quad P = V^{-1} \sim W(R_0, k); \quad V \sim IW(R_0^{-1}, k), \quad (20)$$

where $W(R_0, k)$ is a three-dimensional Wishart distribution with a hyperparameter k ($k \geq 3$ for the current model), a 3×3 *positive definite matrix* R_0 , and the expectation of kR_0 , whereas $IW(R_0^{-1}, k)$

is the corresponding *Inverse Wishart* distribution whose expectation is $R_0^{-1}/(k - 3 - 1)$. Both R_0 and R_0^{-1} must be positive definite; this leads to $k > 4$. When k is lower, the distribution is less informative [28]. Since we have no prior knowledge about V or P , we consider *noninformative prior*. Thus, we set $k = 5$ in this work, and we have $E[V] = R_0^{-1}$.

On the basis of the preceding discussion on the prior, we set

$$R_0^{-1} = \begin{pmatrix} 10.0 & 1.0 & 1.0 \\ 1.0 & 10.0 & 1.0 \\ 1.0 & 1.0 & 10.0 \end{pmatrix} \quad R_0 = \begin{pmatrix} 0.101851852 & -0.009259259 & -0.009259259 \\ -0.009259259 & 0.101851852 & -0.009259259 \\ -0.009259259 & -0.009259259 & 0.101851852 \end{pmatrix}. \quad (21)$$

We can confirm that both R_0 and R_0^{-1} are positive definite by calculating all determinants of the three submatrices (i.e., 1×1 , 2×2 , 3×3 matrices with the same (1,1) element as the original one) to be positive.

The standardized loading coefficients can be calculated in a similar way to that of 3C-SEM as $\lambda_{s,jn} = \lambda_{jn}\sigma_n$, $j = 1, \dots, 6$ for $n = 1$, $j = 7, 8, 9$ for $n = 2$, and $j = 10, \dots, 15$ for $n = 3$, where σ_1, σ_2 , and σ_3 are marginal standard deviations of f_{i1}, f_{i2} , and f_{i3} , respectively, which are free parameters to be estimated.

5.3 1C-CFA Model Based on the UV Method

We can get the whole model of 1C-CFA by replacing the construct model of Figure 2(a) with the bottom panel in Figure 2(b). Since this model has a single factor (construct) of the variable f_i , just setting its variance to 1 provides the standardized coefficients because we have employed the logistic function with a unity variance. We thus apply the UV method.

The distribution of f_i is assumed to be normal with a zero mean: $f_i \sim N(0, 1)$.

The measurement equations are formulated as

$$p_{ij}(k) = 1 / \left[1 + \exp \left\{ - \left(\pi / \sqrt{3} \right) \left(\kappa_{jk} - \mu_{ij} \right) \right\} \right]; \quad k = 1, \dots, 4 \quad (22)$$

$$\mu_{ij} = \lambda_j \cdot f_i; \quad j = 1, \dots, 15. \quad (23)$$

When using the UV method, we must take into consideration the orientation issue of constructs. To avoid the label switching of the loading coefficients, the coefficients are forced to be positive by setting $\lambda_j \sim N(1, 1) \cdot I(0, \infty)$ ($j = 1, \dots, 15$), where $I(0, \infty)$ is the *interval censoring function* [28] that rejects negative values generated by the $N(1, 1)$ distribution. As a matter of fact, unless $I(0, \infty)$ is used, we can observe the label switching.

The priors of the cut points are the same as those of 3C-SEM.

5.4 Model Checking and Comparison

Model checking is to investigate whether the model fit is adequate or not [15, 16, 28]; it examines the consistency or alignment between the data and the model [14].

This work adopts *posterior predictive model checking* (PPMC) using discrepancy functions [14, 28]. We select the discrepancy function to be a *Pearson chi-squared statistic* as in Tasaka [30], although it is “using the data twice” [16] for simplicity. Note that a stimulus in Tasaka [30] corresponds to an observed variable in this article. Letting $D[m]$ denote the discrepancy function for the observed variables z_{im} ($i = 1, \dots, N$) and letting $D_{pred}[m]$ be the one for the corresponding predicted values, we define the *PPp-value* for the m -th observed variable $p_{Bayes}[m]$ as

$$p_{Bayes}[m] \triangleq \Pr(D_{pred}[m] \geq D[m]); \quad m = 1, \dots, 15 \quad (24)$$

and the *overall PPp-value* as the expectation of $p_{Bayes}[m]$ with respect to m : $p_{Bayes} \triangleq E[p_{Bayes}[m]]$.

Fig. 3. Convergence check of qoe[14] (overall satisfaction) for $B_{av} = 100$ ms in 3C-SEM and 3C-CFA models for the castanets hitting.

Note that a p_{Bayes} value around 0.5 implies a plausible model, whereas values toward the extremes of 0 or 1 indicate lack of fit; mid-range values (between 0.2 and 0.8) indicate satisfactory models [16].

For model comparison, we utilize DIC [15, 16, 28], which is the most popular measure of fit in the MCMC simulation. The software OpenBUGS can calculate the DIC. A model with a smaller value of DIC is regarded as more plausible. Only differences in DIC across models are meaningful; absolute values of DIC have no intrinsic meaning.

Regarding the DIC difference, Lunn et al. [28] mention that differences of more than 10 might definitely rule out the model with the higher DIC, and differences between 5 and 10 are substantial.

6 SIMULATION RESULTS

We now present numerical results obtained by MCMC simulation of the 3C-SEM, 3C-CFA, and 1C-CFA models with the software OpenBUGS for both castanets hitting and object movement, and perform the model checking and comparison. We then investigate causation between the constructs and distinguish it from the correlation.

For the castanets hitting, we assessed six stimuli, which correspond to the six values of the audio-video buffering time B_{av} (20, 40, 60, 100, 150, and 300 ms) and five stimuli ($B_{av} = 20, 40, 60, 100$, and 150 ms) for the object movement.

In all 3C-SEM models and the 3C-CFA model with $B_{av} = 20$ ms, we ran three chains for an 80,000 iteration period following a 1,000-iteration *burn-in*; *thinning* was set to 10. The remaining 3C-CFA models employed three chains for a 48,000 iteration period with the thinning of 5 after a 3,000 burn-in, and for all 1C-CFA models, we set three chains for a 50,000 iteration period with the thinning of 5 after a 1,000 burn-in. In each simulation run, the *Brooks–Gelman–Rubin* (BGR) *diagnostic* $\hat{R}$ of every parameter has achieved values very close to 1, by which we can confirm the convergence of the simulation run.

6.1 Model Checking and Comparison

As an example, Figure 3 shows the BGR diagnostic, which is represented by the top line among the three, autocorrelation function, and posterior density of qoe[14] (QoE for overall satisfaction) for the castanets hitting when $B_{av} = 100$ ms in the 3C-SEM (top) and 3C-CFA (bottom) models; this indicates the convergence of the simulation runs.

Fig. 4. Posterior means of QoE measures, qoe[1], qoe[7], and qoe[14], for the task of castanets hitting.

Fig. 5. DIC and PPp-value (p_{Bayes}) versus B_{av} for the task of castanets hitting.

In Figure 4, we plot posterior means of QoE measures qoe[1] (Video Temporal quality), qoe[7] (Haptic Smoothness), and qoe[14] (Overall Satisfaction) for the castanets hitting versus B_{av} for 1C-CFA, 3C-CFA, and 3C-SEM models, along with the 95% credible intervals; for comparison purposes, the corresponding MOS values with 95% confidence intervals are displayed. We then see that the four kinds of QoE measures are very close to each other. This is the case with the other QoE measures, as well as those for the object movement.

We now compare the 3C-SEM, 3C-CFA, and 1C-CFA models in terms of the DIC value and PPp-value (p_{Bayes}) to find the most plausible model. Figures 5 and 6 plot the DIC value and PPp-value as a function of B_{av} for the castanets hitting and object movement, respectively. From these two figures, we see that 3C-SEM has the minimum DIC value for each B_{av} and that the differences in DIC between 3C-SEM and another model (3C-CFA or 1C-CFA) is more than 10, which supports superiority of 3C-SEM over the other two models. We also notice that the PPp-values (p_{Bayes}) exhibit satisfactory fit since they are in a range of [0.43, 0.49]. We have thus learned that 3C-SEM is the most plausible among the three models, and therefore the causation between AVQ and UXQ, as well as that between HQ and UXQ, is more plausible than the corresponding correlations.

Owing to this finding, we focus only on 3C-SEM's results in the following discussion.

6.2 QoE Estimates by 3C-SEM

Figure 7 shows posterior means of the QoE measures for the three constructs as a function of B_{av} for the castanets hitting (top three panels) and the object movement (bottom three panels).

Fig. 6. DIC and PPp-value (p_{Bayes}) versus B_{av} for the task of object movement.

Fig. 7. Posterior means of observed variables versus audiovisual buffering time for three constructs.

In Figure 7, we first notice that the QoE measures of the castanets hitting are more sensitive to the B_{av} value than those of the object movement; each of them has a clear peak in the castanets hitting. This implies that longer buffering time degrades QoE owing to the requirement for fast response in the castanets hitting and the tradeoff relationship between fidelity and latency in interactive communications [30].

In addition, it should be noted that the QoE curves even for HQ have a clear peak in the castanets hitting. Recall that the traffic conditions at the application layer for HQ are kept constant irrespective of the B_{av} values because B_h is fixed to 10 ms, whereas the conditions for AVQ change

Fig. 8. Comparison of posterior means of qoe[15]/qoe[14] and the average number of completed tasks.

as B_{av} increases. This suggests that the acuminate curves for HQ are generated by the effect of AVQ. We can regard it as a sort of *inter-modality phenomenon* or *cross-modal effect* between haptic and visual media. This phenomenon is also observed in the object movement, although the change is gradual compared to that in the castanets hitting. The cross-modal effects between audio and video have been studied in many publications (e.g., [32, 33]). However, it is difficult to find some systematic research on the effects between haptic and visual media in the literature.

We next examine the accuracy of the approximation made for n_c ($NCmpl$) about the truncation to 5 in Section 4.1; n_c ($NCmpl$) denotes the number of instructions given by the instructor in the castanets hitting and the number of objects the manipulator completed the movement to the destination in the object movement during an experimental run. Figure 8 shows the posterior means of qoe[15] for the castanets hitting and qoe[14] for the object movement along with the corresponding average number of $NCmpl$ (i.e., without the approximation) as a function of B_{av} . We notice that the QoE measure for the castanets hitting, qoe[15], is a little bit smaller than the average of $NCmpl$, whereas the two kinds of the values are close to each other in the object movement; this is because n_c never exceeded 5 in the experiment of the object movement.

6.3 Causation and Correlation Between Constructs

We are now in a position to investigate the causation and correlation issue in 3C-SEM, which is the core of this article.

Before we study the issue, look at Figure 9, which displays posterior densities of correlation coefficients between the constructs for 3C-SEM and 3C-CFA in the castanets hitting, and we notice the distributions being skewed; this is also the case with the object movement. Therefore, we do not use posterior means but present posterior *medians* of causation measures and correlation coefficients.

The left panel of Figure 10 plots posterior medians of the unstandardized regression coefficients (γ_1, γ_2) and the standardized regression coefficients ($\gamma_{s,1}, \gamma_{s,2}$) versus B_{av} for the castanets hitting in 3C-SEM. Comparing $\gamma_{s,1}$ and $\gamma_{s,2}$, we immediately see that the causal effect of HQ on UXQ is much stronger than that of AVQ on UXQ when B_{av} is large. The right panel of the figure depicts the correlation coefficients $\rho_{13}^{(SEM)}$ and $\rho_{23}^{(SEM)}$ along with $\gamma_{s,1}$ and $\gamma_{s,2}$ by using Equations (18) and (19). From this panel, we learn that although the correlation between AVQ and UXQ is middle to high, the contribution of the causal effect is small. Regarding the relationship between HQ and

(a) Posterior density of correlation coefficient for 3C-SEM (castanets, $B_{av} = 100$ ms) (b) Posterior density of correlation coefficient for 3C-CFA (castanets, $B_{av} = 100$ ms)

Fig. 9. Posterior densities of correlation coefficients for 3C-SEM and 3C-CFA: castanets hitting.

Fig. 10. Posterior medians of regression coefficients and correlation coefficients for 3C-SEM: castanets hitting.

UXQ, however, the difference between the causal effect and the correlation coefficient is small; the correlation is mainly due to the causation.

Figure 11 shows the posterior medians for the object movement in the same way as that of the castanets hitting. Comparing Figures 10 and 11, we first find that the object movement yields stronger causal effects of HQ on UXQ than those of AVQ on UXQ for all of the B_{av} values and that the correlation coefficient contributed by the causal effects has similar characteristics to those of the castanets hitting.

For more information on the correlation coefficients, Figure 12 demonstrates three correlation coefficients, ρ_{12} , ρ_{13} , and ρ_{23} , for both 3C-CFA and 3C-SEM in the castanets hitting (left) and object movement (right). We observe the same tendency of the correlation coefficients according to the increase in B_{av} in the two tasks, but the difference of ρ_{23} between 3C-CFA and 3C-SEM in the

Fig. 11. Posterior medians of regression coefficients and correlation coefficients for 3C-SEM: object movement.

Fig. 12. Posterior medians of correlation coefficients: comparison between 3C-CFA and 3C-SEM.

object movement is smaller than that in the castanets hitting, whereas the situations of ρ_{12} and ρ_{13} are rather opposite.

From the findings so far, we learn that the relationships of causation and correlation depend on the kind of task, but stronger causal effects of HQ on UXQ are common. This suggests to us that the improvement of haptic performance can contribute to enhancement of QoE more effectively than that of audiovisual performance.

6.4 Standardized Loading Coefficients in 3C-SEM

Finally, we examine how the indicators load on the constructs in 3C-SEM. Figure 13 shows box plots of the standardized loading coefficients of 3C-SEM with $B_{av} = 60$ and 100 ms for the castanets hitting (left) and for the object movement (right).

From this figure, we see that the loading coefficients depend on the B_{av} value and the task, and the values of the coefficients $\lambda_{s,15,3}$ for the castanets hitting and $\lambda_{s,14,3}$ for the object movement,

Fig. 13. Posterior means of standardized loading coefficients for 3C-SEM.

which correspond to $NCmpl$, are small compared to the others; this is natural because $NCmpl$ is an objective measure.

We also observe that in the castanets hitting, $\lambda_{s,5,1}$ (Audio Naturalness) and $\lambda_{s,6,1}$ (Castanet Sync) are small, and that $\lambda_{s,5,1}$ (Audio Naturalness) in the object movement with $B_{av} = 100$ ms is larger than the average of the loading coefficients.

7 CONCLUSION

This article proposed a methodology for studying the causation and correlation in haptic-audiovisual interactive communications by means of Bayesian SEM and CFA. The SEM and CFA were utilized for dimension reduction of multidimensional QoE by introducing latent variables (constructs and score continuous latent variables). At its cost, we had to treat the identifiability problem of latent variables and further calculate standardized coefficients of regression and loading for fair comparison. We employed two tasks (castanets hitting and object movement) and supposed three constructs (AVQ, HQ, and UXQ) and a single construct for each task. We then built SEM and CFA models each with three constructs (3C-SEM and 3C-CFA) and a CFA model with a single construct (1C-CFA). Comparing the three models for each task, we found that 3C-SEM is the most plausible among the three; therefore, the causal effect of AVQ and UXQ and that of HQ on UXQ are more plausible than only the correlations between them, and that the latter causal effect

is much stronger than the former. We also noticed that the correlation between AVQ and UXQ is much larger than the causal effect between the two; the difference between the correlation and the causal effect is produced by the causal effect of HQ on UXQ via correlation of AVQ with HQ. This suggests that improving haptic performance is more effective in enhancement of QoE than improving audiovisual performance.

It should be emphasized again that 3C-SEM is not necessarily the best solution to the causation-correlation problem treated in this article but a plausible solution in comparison to the 3C-CFA and 1C-CFA models. We could construct many variants of the models proposed in this article but did not do so, as this work aimed at establishing a methodology for the problem.

The proposed methodology can be extended in several ways. To improve the models and extend the applicability, we can utilize physiological data and objective data measured in the application and lower layers in addition to subjective data. It is desirable to try to determine causal directions of constructs mainly with subjective data. Future work also includes another approach to the causation-correlation problem in an opposite way to this work; causal relationships between individual QoE measures are first explored, and then reducing the dimension of the relationships is carried out by extracting common features of individual causal relationships.

ACKNOWLEDGMENTS

The author thanks Eiichi Isomura, who provided the experimental data, and Professor Toshiro Nunome of Nagoya Institute of Technology for his discussion on the experiment.

REFERENCES

- [1] Judea Pearl. 2009. *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
- [2] Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell. 2016. *Causal Inference in Statistics, A Primer*. John Wiley & Sons.
- [3] Stanley A. Mulaik. 2009. *Linear Causal Modeling with Structural Equations*. CRC Press, Boca Raton, FL.
- [4] Peter Brooks and Bjørn Hestnes. 2010. User measures of quality of experience: Why being objective and quantitative is important. *IEEE Network* 24, 2 (March/April 2010), 8–13.
- [5] Khalil ur Rehman Laghari, Noel Crespi, and Kay Connelly. 2012. Toward total quality of experience: A QoE model in a communication ecosystem. *IEEE Communications Magazine* 50, 4 (Apr. 2012), 58–65.
- [6] Patrick Le Callet, Sebastian Möller, and Andrew Perkins eds. 2013. *Qualinet White Paper on Definitions of Quality of Experience*, Version 1.2 (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland.
- [7] Eckehard Steinbach, Sandra Hirche, Marc Ernst, Fernanda Brandi, Rahul Chaudhari, Julius Kammerl, and Iason Vitorias. 2012. Haptic communications. *Proceedings of the IEEE* 100, 4 (April 2012), 937–956. DOI: [10.1109/JPROC.2011.2182100](https://doi.org/10.1109/JPROC.2011.2182100)
- [8] Eiichi Isomura, Shuji Tasaka, and Toshiro Nunome. 2013. QoE enhancement by media adaptive intra-stream synchronization in audiovisual and haptic IP communications. *IEICE Transactions on Communications (in Japanese)* J96–B, 2 (Feb. 2013), 59–70.
- [9] Kenneth A. Bollen. 1989. *Structural Equations with Latent Variables*. John Wiley & Sons.
- [10] Xin-Yuan Song and Sik-Yum Lee. 2012. *Basic and Advanced Bayesian Structural Equation Modeling with Application in the Medical and Behavioral Sciences*. John Wiley & Sons.
- [11] David Bartholomew, Martin Knott, and Irini Moustaki. 2011. *Latent Variable Models and Factor Analysis: A Unified Approach* (3rd ed.). John Wiley & Sons.
- [12] Rex B. Kline. 2016. *Principles and Practice of Structural Equation Modeling* (4th ed.). Guilford Press.
- [13] Karl G. Jöreskog and Dag Sörbom. 1993. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International.
- [14] Roy Levy and Robert J. Mislevy. 2016. *Bayesian Psychometric Modeling*. CRC Press, Boca Raton, FL.
- [15] Peter Congdon. 2006. *Bayesian Statistical Modelling* (2nd ed.). John Wiley & Sons.
- [16] Peter Congdon. 2014. *Applied Bayesian Modelling* (2nd ed.). John Wiley & Sons.
- [17] Shuji Tasaka. 2016. Bayesian structural equation modeling of multidimensional QoE in haptic-audiovisual interactive communications. In *Proceedings of the 2016 International Conference on Communications*. 3345–3350. DOI: [10.1109/ICC.2016.7511202](https://doi.org/10.1109/ICC.2016.7511202)

- [18] Tatsuya Yamazaki, Masato Eguchi, Takumi Miyoshi, and Kyoko Yamori. 2012. A service quality coordination model bridging QoS and QoE. In *Proceedings of the 2012 IEEE 20th International Workshop on Quality of Service (IWQoS'12)*.
- [19] Shuji Tasaka. 2019. Bayesian categorical modeling of multidimensional QoE in haptic-audiovisual communications. In *Proceedings of the 2019 International Conference on Communications*. 1–7. DOI: [10.1109/ICC.2019.8761784](https://doi.org/10.1109/ICC.2019.8761784)
- [20] Abdelwahab Hamam, Nicholas D. Georganas, and Abdulmotaleb El Saddik. 2010. Effect of haptics on the Quality of Experience. In *Proceedings of the 2010 IEEE International Symposium on Haptic Audio Visual Environments and Games (HAVE'10)*. 1–6. DOI: [10.1109/HAVE.2010.5623992](https://doi.org/10.1109/HAVE.2010.5623992)
- [21] Abdelwahab Hamam, Mohamad Eid, and Abdulmotaleb El Saddik. 2013. Effect of kinesthetic and tactile haptic feedback on quality of experience of edutainment applications. *Multimedia Tools and Applications* 67, (2013), 455–472. DOI: [10.1007/s11042-012-0990-7](https://doi.org/10.1007/s11042-012-0990-7)
- [22] Abdelwahab Hamam, Abdulmotaleb El Saddik, and Jihad Alja'am. 2014. A quality of experience model for haptic virtual environments. *ACM Transactions on Multimedia Computing, Communications, and Applications* 10, 3 (April 2014), Article 28, 23 pages. DOI: [10.1145/2540991](https://doi.org/10.1145/2540991).
- [23] Kai Iwata, Yutaka Ishibashi, Norishige Fukushima, and Shinji Sugawara. 2010. QoE assessment in haptic media, sound, and video transmission: Effect of playout buffering control. *Computers in Entertainment* 8, 2 (Dec. 2010), Article 12. DOI: [10.1145/1899687.1899694](https://doi.org/10.1145/1899687.1899694)
- [24] Ayano Tatematsu, Yutaka Ishibashi, Norishige Fukushima, and Shinji Sugawara. 2010. QoE assessment in haptic media, sound, and video transmission: Effect of playout buffering control. In *Proc. 2010 IEEE Intern. Workshop Tech. Committee on Commun. Quality Reliability (CQR2010)* 6 pages, June 2010. DOI: [10.1109/CQR.2010.5619913](https://doi.org/10.1109/CQR.2010.5619913)
- [25] Eiichi Isomura, Shuji Tasaka, and Toshiro Nunome. 2013. A multidimensional QoE monitoring system for audiovisual and haptic interactive IP communications. In *Proceedings of the 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC'13)*. 196–202.
- [26] J. Kenneth Salisbury and Mandayam A. Srinivasan. 1997. Phantom-based haptic interaction with virtual objects. *IEEE Computer Graphics and Applications* 17, 5 (Sept./Oct. 1997), 6–10. DOI: [10.1109/MCG.1997.10014](https://doi.org/10.1109/MCG.1997.10014)
- [27] James L. Arbuckle. 2012. *IBM SPSS Amos21 User's Guide*. Amos Development Corporation.
- [28] David Lunn, Christopher Jackson, Nicky Best, Andrew Thomas, and David Spiegelhalter. 2013. *The BUGS Book*. CRC Press, Boca Raton, FL.
- [29] OpenBUGS. 2015. OpenBUGS License. Retrieved January 27, 2020 from <http://www.openbugs.net/w/Downloads>.
- [30] Shuji Tasaka. 2017. Bayesian hierarchical regression models for QoE estimation and prediction in audiovisual communications. *IEEE Transactions on Multimedia* 19, 6 (June 2017), 1195–1208. DOI: [10.1109/TMM.2017.2652064](https://doi.org/10.1109/TMM.2017.2652064)
- [31] John C. Loehlin and A. Alexander Beaujen. 2017. *Latent Variable Models*. Routledge.
- [32] Harry McGurk and John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264 (Dec. 1976), 746–748.
- [33] Shuji Tasaka and Yoshihiro Ito. 2003. Psychometric analysis of the mutually compensatory property of multimedia QoS. In *Proceedings of the IEEE International Conference on Communications*. 1880–1886.

Received August 2019; revised December 2019; accepted December 2019