Customer Decision-Making Processes Revisited: Insights from an Eye Tracking and ECG Study using a Hidden Markov Model

Tobias Weiß, Lukas Merkl, and Jella Pfeiffer

Justus Liebig University Gießen, Gießen, Germany
{tobias.weiss,lukas.merkl,jella.pfeiffer}@wirtschaft.uni-giessen.de

Abstract. Good timing is key for many activities in business and society. In the context of adaptive user assistance, it can work as door opener to further engage with the user. This paper presents a virtual commerce study which combines eye tracking, electrocardiography, and virtual reality with the goal to detect decision phases in two different purchase scenarios. We therefore collect objective sensor data in combination with subjective decision phase annotations. Shifts between decision phases are determined subjectively by the participants via retrospective video analysis. For decision phase recognition, we demonstrate how to use the neurophysiological sensor data to train a Hidden Markov Model with multivariate mixed Gaussian emission distributions and how to use it for inference. A main benefit of our approach is its lightweight character regarding both training and inference.

Keywords: Customer Behavior, Decision Making, Eye Tracking, Electrocardiography, Hidden Markov Model, Gaussian Mixture Model, Machine Learning, Virtual Commerce, Virtual Reality.

1 Introduction

Approaching customers at the right time is crucial because it can significantly impact the interaction success [27]. Specifically, good timing can help to maximize engagement, build trust, and increase conversion rates [6]. However, to determine the right point in time to approach a customer requires profound understanding of the target audience's behavior and preferences [8]. Advances in conversational agents and user assistance systems (UAS) often focus on the right information, introduce context awareness and improve interactivity [12, 17, 27]. Rather scarcely, previous research has investigated invocation timing based on neurophysiological indicators [16]. Within the ongoing transformation of the retail sector towards virtual commerce [3] and the rise of the metaverse idea [1], good invocation timing is one of the key components for a variety of information system (IS) artifacts. Decades ago, metaverse and virtual reality (VR) advocates already envisioned that a large fraction of daily life and therewith a large fraction of shopping activities transfers to virtual spaces [10, 25]. Today, this process gains momentum, as big tech companies introduce new hardware and applications with rigorous commitment. Latest VR headsets ship with eye and face tracking technology which fosters the potential and feasibility of neurophysiological IS and therefore turns them into a game changer. With a VR headset on their head, future customers wear a variety of sensors in proximity to the most reliable information source about their attitudes and moods. In this paper, we present our approach to integrate neuroscientific methods into virtual commerce IS. Our research question states as follows:

RQ: Can we determine a good timing to approach customers in a virtual commerce scenario using eye-tracking and electrocardiography?

We report our insights gained from a study in which 50 participants had to make purchase decisions for either washing powder or 3D printers while wearing a head-mounted VR headset. We collected participants' eye tracking (ET) data, electrocardiography (ECG) data, and created a prediction model that can distinguish between different decision phases. Our insight can be used to inform a UAS or digital human agent when help is wanted. As model for decision phase recognition, we chose a combination of multivariate Gaussian Mixed Model and Hidden Markov Model (GMM-HMM). The benefit of our approach is its lightweight character in both training and inference. Thus, the presented GMM-HMM approach offers itself as good candidate to make it into soon-to-be released virtual commerce (and other) neurophysiological sensor based IS artifacts [28]. To the best of our knowledge, no study exists which applies machine learning approaches to differentiate between different decision phases using neurophysiological sensor data. Our research builds up on previous models but tries to apply a more generic inference method not solely dependent on product comparisons and re-dwells.

2 Method

2.1 Consumer Decision-Making

Several scholars in consumer behavior research suggested models to subdivide customer decision-making processes into different phases. Most studies support a phase theory which consists at least of an orientation and an evaluation phase. One prominent phase model is the five-stage Engel Kollat Blackwell (EKB) model [5], as shown in Figure 1. The EKB model is still widely accepted [24] and frequently serves as basis for further adjustment to integrate specific aspects and research field dependent needs, such as modifications for an eye tracking study in VR.

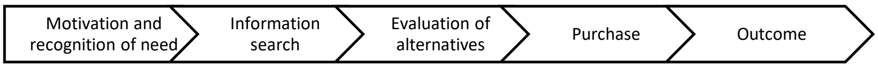

Fig. 1. The EKB model, dividing customer decision processes into five phases [5].

In an eye tracking context, several other decision phase models were developed, e.g., by [21], [7], and [16]. These models subdivide decision processes into three phases – orientation, evaluation, and validation. The transition between different phases is based on simple rules, like re-fixations on products. The VR study in [16] pursued an on-the-fly attempt to determine the phases. Its authors used eye tracking data and identified the first comparison between two products as shift between orientation and evaluation. Furthermore, the shift between evaluation and verification was considered as the moment when the first product entered the shopping cart (We believe this is a questionable criterion because putting a product into the shopping cart signals a certain level of confidence).

For the right timing of user assistance, we consider the shift between orientation and evaluation as particularly interesting. We conjecture that help is most appreciated by customers after being within the evaluation phase for a certain offset duration. To verify this assumption empirically, self-reported desired help timings can be used. Knowing the phase of a decision process and the offset duration at least approximately, a UAS or sales representative can determine a good starting point to approach the customer.

2.2 Neurophysiological Data Collection in VR

The development of visual VR has a longer history than one might expect. For example, an early head mounted display (HMD) was already developed by Sutherland [26]. Commercial endeavors of big tech companies still focus on HMD development. For research, the latest HMD generation is particularly interesting because many models ship with integrated neurophysiological sensors, particularly ET [18]. ET is integrated because it can be used to optimize graphic card utilization via foveated rendering, a method which only renders the focused area in high detail while neglecting peripheral areas [14]. Recent research-grade HMDs include further sensors as ECG and electroencephalography (EEG). The integration of EEG into consumer-grade hardware seems rather unrealistic in the near and intermediate-term future as the sensor itself is expensive and the electrodes are relatively uncomfortable to wear. ECG measures a person's heart rate and is more likely to find its way into consumer devices. Another sensor, which is very likely to be included into future customer-grade HMDs, is photoplethysmography (PPG). PPG is a light-based sensor which can also be used to measure heart rate and corresponding metrics. Compared to ECG, PPG is cheaper, easier to attach (e.g., a forehead-sensor integrated in the HMD-cover), but less accurate. It is also imaginable to couple wearables with an HMD, particularly fitness watches, which already include ECG or PPG sensors. Overall, ET and ECG/PPG are the most likely sensors for future off-the-shelf HMDs. Thus, it makes sense to use gaze patterns, pupillometry, and heart rate as data sources for inference.

2.3 Hidden Markov Model

An HMM is a statistical model which describes a Markov process with a set of states between which it can transition [4, 20]. At each state, an HMM generates an observation or output symbol, which is associated with that state. Such observations generated by a state of the model are referred to as emissions. HMMs find application in a variety of disciplines [9, 11, 23]. To match the characteristics of our purchase decision scenario in the experimental VR setup, we use elements of both the classic EKB phase model [5] and the eye tracking model proposed by Russo et al. [21]. We begin with a memorization phase which corresponds to the motivation phase of the EKB model. During this phase, participants see purchase criteria on a blackboard and memorize them. The transition between memorization and the next phase is identified by a button press. For the subsequent phases, we use the phase labels orientation, evaluation and verification as proposed by Russo et al. [21]. However, we outline that the state transitions in our model have nothing in common with the originally proposed transitions which were based on specific gaze patterns. Instead, shifts to evaluation and verification were determined via self-reported timestamps given by the participants. Next, we adopt the purchase phase from the EKB model, as participants remained inside the VR scenario after confirming the purchase. Furthermore, an initial and terminal state are added as they are needed for computation. The corresponding HMM with flat prior transition probabilities is shown in Figure 2.

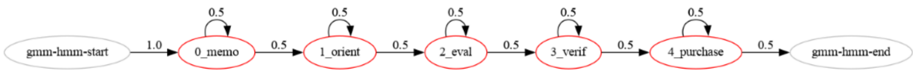

Fig. 2. GMM-HMM with flat prior transition probabilities.

When the model transitions from one state to another, it refers to a (hidden) multivariate probability distribution which corresponds to the current input features. Internally, each state holds a multivariate Gaussian mixture distribution (what turns the model into a GMM-HMM), which is trained with ET and ECG features based on consecutive five second time windows. For each of these windows, our feature engineering pipeline creates 44 features which comprise 26 ET and 18 ECG features. ET features consist of statistical moments (mean, min, max, var) for blinks, fixations, fixation duration, pupil size, saccadic duration, and saccadic speed. ECG features are limited to the time domain, particularly the heart rate and its variability. Frequency domain related and non-linear ECG features are not considered because they would require longer window durations [19]. If participants indicate a state transition during such a window, the label for the subsequent and all following windows changes to the next state.

For real-time inference, the GMM-HMM can even be simplified to a GMM classifier which decides if the evaluation phase is reached or not. Features of a current observation can be shown to the model which maps them to the probability distribution and stochastically decides whether the evaluation phase is reached or not. If the evaluation phase is indicated several times in a row, the offset of approximately fifty seconds could be added and finally the UAS or digital human agent could approach the customer with a help offering.

3 Experiment

3.1 Participants

Our sample was collected from 50 participants (29 female, mostly students) with a mean age of 24.5 years (SD=4.89). Only individuals with normal or corrected-to-normal vision via contact lenses were accepted since glasses would not fit into the HMD and not wearing them might confound the ET data. The participation compensation consisted of a fixed 10 Euro baseline plus a performance-based component. After arrival at the lab, participants signed a consent form. It ensured the participants' basic knowledge of the experiment procedure and informed them that the experiment complied with ethical standards. Further, it required them to grant permission to publish their data in anonymized form. For recruitment, we used the participant pool in our self-hosted online registration platform [2] and actively approached students on campus.

3.2 Procedure

We simulated customer purchase decisions in VR, collecting ET and ECG data. All virtual scenes were created using the Unity 2021.3 game engine. Participants entered our showroom using a Varjo VR 3 HMD with high-frequency ET capability (sampling rate of up to 200 Hz) and a display resolution of 2880 × 2720 pixels per eye. A bio-PLUX device was used for ECG recording and captured signals with a sampling rate of 1000 Hz. Overall, the experiment followed a between-subjects design and included two different decision scenarios, one for 3D printers and one for washing powders (see Figure 3). To create realistic shopping scenarios, we presented dedicated cover stories to both groups. Participants were then shown a list of purchase decision criteria they had to memorize. The end of memorization phase was triggered by the participants using a button press which hid the criteria and spawned the products. Then, they had the chance to gain one Euro in addition to their participation compensation if they matched a previously determined team decision. This monetary incentive helped to motivate the participants and increased the external validity of the experiment. Participants confirmed their purchase decision either by putting the product into a shopping cart or by clicking a purchase button. After making the purchase, participants left the VR environment and answered questions about their decision phases by means of a first-person video. This video showed a gaze dot which indicated their visual attention. Participants determined the moments when they shifted (1) from orientation to evaluation and (2) from evaluation to verification. For each of these phase shifts, they entered the timestamp in a web-based questionnaire form. Furthermore, participants reported their desired help timing for a digital human agent in the same manner as for the phase shifts.

Fig. 3. Experimental VR setup (3D printer decision top, washing powder decision bottom).

4 Results

For our analysis, we used python 3.10 and the neurokit2 0.2.3 [13], pomegranate 0.4.0 [22], and scikit-learn 1.0.2 [15] packages.

We verified our conjecture regarding the desired help timing. As expected, help was most frequently desired after the shift from orientation to evaluation but before entering the verification phase. On average, the phase shift from orientation to evaluation was indicated after 100.2 seconds (SD=79.8) and the shift from evaluation to verification was after 210 seconds (SD=97.2). Participants reported the average desired help timing for a digital human agent after 148 seconds (SD=115.8), i.e., with an average offset of 48 seconds after starting the evaluation phase and 62 seconds before entering the verification phase (see Figure 4).

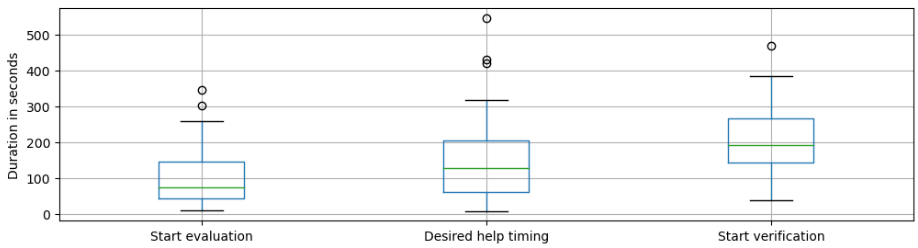

Fig. 4. Boxplots of the self-reported phase shifts and the desired help timing.

Our trained model with posterior transition probabilities is shown in Figure 5. Each state is holding a multivariate GMM which consists of multiple Gaussian mixture distributions (see Figure 6 left for a univariate example). We showcase the inference of one full exemplary purchase process in Figure 6 right. Such phase predictions can be further refined and leveraged by a UAS or sales agent to find the best time to approach customers with an assistance offering. It is noteworthy that training duration only lasted 3.21 seconds and with very brief inference times a single observation can be predicted on the fly. The mean difference between classified and reported shifts from orientation to evaluation is -0.14 (SD=4.49) five second time windows. Overall, the model fits 84.89% of the five second windows correctly.

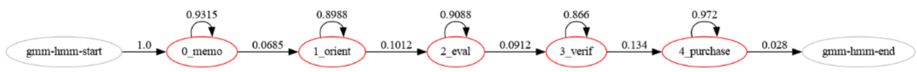

Fig. 5. GMM-HMM with posterior transition probabilities.

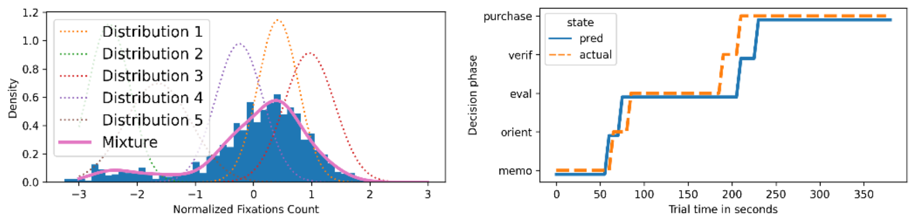

Fig. 6. Exemplified univariate GMM for a single feature (left), comparison between reported state transitions and model prediction for one purchase decision (right).

5 Discussion

Our results show the feasibility of identifying a good timing to approach customers in a virtual commerce scenario using GMM-HMMs and thus yield an answer to our research question. The presented approach uses multiple neurophysiological sensors as input and meets our goal to overcome pure comparison and fixation-based phase determination. The presented methodology can be adopted by other researchers and practitioners to build a maybe soon to be realized overarching virtual platform, offering a multitude of interconnected virtual worlds and services.

This work has limitations which may serve as a guideline for future research. First, our sample almost exclusively consists of students, which limits generalizability. Future research should involve a broader cross-section of society. Second, the sample size should be increased. Our 50 observations yield little variety to equip the model with performant predictive power. Third, immersion, perceived telepresence, and perceived product involvement could have been increased by adding more sensory channels (particularly audio) to the virtual environment. Room size also played a limiting role, as participants had to remain relatively immobile and could not fully immerse themselves in the virtual space. Regarding the applied machine learning techniques, we plan to rigidly quantify the model performance and give detailed information about the most relevant features. We also want to consider further measurements as features, such as electrodermal activity and electroencephalography, which eventually might also be integrated into future HMDs off-the-shelf. Finally, we plan to evaluate the simplified GMM classifier version of the model in an experimental virtual commerce shopping scenario in which a digital human agent approaches a customer according to the timing suggested by the model.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/1 – Project Nr. 447089431 – Research Training Group: KD²School – Designing Adaptive Systems for Economic Decisions

References

1. Ball M (2022) The Metaverse. Liveright, [S.I.]
2. Bock O, Baetge I, Nicklisch A (2014) hroot: Hamburg Registration and Organization Online Tool. European Economic Review 71:117–120
3. Bourlakis M, Papagiannidis S, Li F (2009) Retail spatial evolution: paving the way from traditional to metaverse retailing. Electron Commer Res 9:135–148
4. Eddy SR (2004) What is a hidden Markov model? Nature biotechnology 22:1315–1316
5. Engel JF, Kollat DT, Blackwell RD (1968) A model of consumer motivation and behavior. Research in Consumer Behavior, Holt, Rinehart and Winston, Inc., New York, NY:3–20
6. Friemel C, Morana S, Pfeiffer J et al. (2018) On the Role of Users’ Cognitive-Affective States for User Assistance Invocation. In: Information Systems and Neuroscience. Springer, Cham, pp 37–46
7. Gidlöf K, Wallin A, Dewhurst R et al. (2013) Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. JEMR 6
8. Horvitz EJ, Breese JS, Heckerman D et al. (2013) The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. arXiv preprint arXiv:1301.7385
9. Krogh A, Brown M, Mian IS et al. (1994) Hidden Markov models in computational biology: Applications to protein modeling. Journal of molecular biology 235:1501–1531
10. Lanier J, Biocca F (1992) An Insider's View of the Future of Virtual Reality. Journal of Communication 42:150–172
11. Liu H, Gamboa H, Schultz T (2023) Sensor-Based Human Activity and Behavior Research: Where Advanced Sensing and Recognition Technologies Meet. Sensors 23

12. Maedche A, Morana S, Schacht S et al. (2016) Advanced user assistance systems. Bus Inf Syst Eng 58:367–370
13. Makowski D, Pham T, Lau ZJ et al. (2021) NeuroKit2: A Python toolbox for neurophysiological signal processing. Behav Res 53:1689–1696
14. Patney A, Salvi M, Kim J et al. (2016) Towards foveated rendering for gaze-tracked virtual reality. ACM Trans Graph 35:1–12. doi: 10.1145/2980179.2980246
15. Pedregosa F, Varoquaux G, Gramfort A et al. (2011) Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12:2825–2830
16. Peukert C, Lechner J, Pfeiffer J et al. (2020) Intelligent Invocation: Towards Designing Context-Aware User Assistance Systems Based on Real-Time Eye Tracking Data Analysis. In: Information Systems and Neuroscience. Springer, Cham, pp 73–82
17. Pfeiffer J (2011) Interactive decision aids in e-commerce. Springer Science & Business Media
18. Pfeiffer J, Pfeiffer T, Meißner M et al. (2020) Eye-Tracking-Based Classification of Information Search Behavior Using Machine Learning: Evidence from Experiments in Physical Shops and Virtual Reality Shopping Environments. Information Systems Research 31:675–691
19. Pham T, Lau Z, Chen SH et al. (2021) Heart rate variability in psychology: a review of HRV indices and an analysis tutorial. Sensors 21:3998
20. Rabiner L, Juang B (1986) An introduction to hidden Markov models. ieee assp magazine 3:4–16
21. Russo JE, Leclerc F (1994) An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables. J Consum Res 21:274
22. Schreiber J (2017) Pomegranate: fast and flexible probabilistic modeling in python. the Journal of machine Learning research 18:5992--5997
23. Schultz T, Waibel A (2001) Language-independent and language-adaptive acoustic modeling for speech recognition. Speech Communication 35:31–51
24. Sihi D (2018) Home sweet virtual home. JRIM 12:398–417
25. Stephenson N (2003) Snow crash. A Novel. Random House Publishing Group
26. Sutherland IE (1965) The ultimate display. Proceedings of the IFIP Congress 2:506–508
27. Sykes TA (2015) Support Structures and Their Impacts on Employee Outcomes. MIS Quarterly 39:473–496
28. vom Brocke J, Riedl R, Léger P-M (2013) Application strategies for neuroscience in information systems design science research. Journal of Computer Information Systems 53:1–13