Personalized Multimodal Signal Processing for Enhanced HCI in AR Environments

Phivos Mylonas, Christos Troussas, Akrivi Krouska, Cleo Sgouropoulou
Department of Informatics and Computer Engineering
University of West Attica
Egaleo, Greece
{mylonasf, ctrouss, akrouska, csgouro}@uniwa.gr

***Abstract* — This paper presents a new framework of personalized multimodal signal processing specifically tuned into Augmented Reality (AR) environments. Our approach is based upon advanced Machine Learning (ML) algorithms to create adaptive interfaces through the integration of audio, visual, and haptic signals that respond preferentially to individual user traits and behaviors. Our focus is on interactive real-time signal processing with respect to gesture and facial recognition and personalized content delivery in AR applications. The system is evaluated through a user study, showing a significant increase in user engagement and its quality through the mix of interactivity factor when compared to existing non-personalized approaches. The results emphasize the effectiveness of personalized signal processing overhaul toward human-computer interaction (HCI) for the future AR tech developments.**

***Keywords* — personalized signal processing; Augmented Reality; multimodal HCI;real-time content adaptation;**

I. Introduction

The proliferation of Augmented Reality (AR) is placing new parameters around interaction among computers and humans, including different immersive experiences that transverse the physical and digital [1]. However, the success of any AR system is subject to each individual's adjustments, which is paramount in ensuring higher user engagement and satisfaction. With that said, the standard AR systems mostly treat users as homogeneous entities, not considering the different preferences, behaviours, and cognitive loads of individuals [2]. This limitation brings forth the need to suitably approach personalized multimodal signal processing, which combines the many sensory inputs, including those of audio, visual, and haptic signals, to produce adaptive and responsible AR interfaces.

Personalized multimodal signal processing becomes that much more relevant in the AR environments, where individuals interact with virtual objects overlaying the real world [3]. For example, in AR-based training applications for medical students, some possible implementations include adjusting the complexity of visual instructions based on indications expressed by the user, including facial expressions that can denote confusion or comprehension. Again, haptic feedback in games may be adjusted to meet the particular sensitivity to tactile stimuli that each individual should have to immerse themselves in AR gaming. These examples point to revolutionizing what personalization can achieve in AR; however, many challenges still need resolution, such as real-time signal processing, multimodal data fusion, and modelling user behaviours.

This paper addresses these challenges by proposing a broad framework for personalized multimodal signal processing in AR environments. Our approach embodies combining cutting-edge machine learning skills with advanced signal processing algorithms to create adaptive interfaces responding to individual users in real-time. The framework is designed to build upon its modularity to integrate itself smoothly with existing AR systems. Our emphasis on personalization should merge the rift between stereotypical AR interfaces and user-centric experiences; hence, more intuitive human-computer interaction (HCI) in AR business applications will be realized.

The remainder of this paper is organized as follows. Section II reviews related work in multimodal signal processing and personalization in AR. Section III presents the proposed framework, detailing its architecture and key components. Section IV describes the experimental setup and evaluation metrics. Section V discusses the results and their implications for AR applications. Finally, Section VI concludes the paper and outlines future research directions.

II. Related Work

Though it has been a blaze of interest for the last few years due to the ever-increasing demand for immersive interactive experiences, integration of the multi-modal signal processing in AR started focusing more or less on visual signal processing, where earlier techniques used computer vision to track objects and overlay virtual content [6-10]. Nonetheless, multi-modal signal processing started surfacing after AR used instances including healthcare, education, and entertainment [11, 12]. As an example, in AR-based surgery simulation training, exclusively vision-based signals do not offer a sense of realism during the act; haptic feedback along with sound signals could bolster the intensity of training [13].

The increase in ML innovations recently has further sped up the evolution of multimodal signal processing techniques [14]. Exceptional successes with deep architectures have been achieved in the areas of gesture recantation, facial analysis, and emotion recognition [15]. These models would equip AR systems to interpret complex user behaviours and appropriately adapt responses to them [16-19]. For instance, an autonomous virtual assistant in the AR environment can relate the user's tone of voice with their facial expression to determine a user's emotional state and accordingly adapt an interaction style [20]. However, most existing systems treat all users as one and fail to personalize the experience, not considering differences of the individual concerning interests and behaviour.

In different applied perspectives, personalization techniques in AR had been elaborated so far, still very less is done on the integration of multimodal signal processing therein. Very few works focused on either personalizing visual content according to the user his/her preference or adaptive haptic feedback, with no consideration of synergy between various modalities while offering adaptive AR experiences [21-25]. For example, an AR set-up involving personalization of visual content but lone in ignoring the auditory or haptic signals will not deliver orderly user experience [26]. This obviously raises the necessity for an integrated modality framework for successful tailoring of an AR experience to individual users.

The proposed personalized multimodal signal processing framework for AR would lay down the guidelines and protocols to circumvent the hurdles faced by the existing approaches. By combining advanced machine learning techniques with real-time signal processing, we aim to create AR interfaces that are immersive, adaptive, and user-centric. This framework marks a major step forward in AR research, opening up new possibilities for improving human-computer interaction across a wide range of applications.

III. Proposed Framework

The proposed framework for personalizing multimodal signal processing in augmented reality environments tends toward the issues of real-time adaptability and user-centricity. The proposed framework comprises three main components: a multimodal signal acquisition module, a personalization engine, and an adaptive interface. Fig. 1 illustrates the proposed framework.

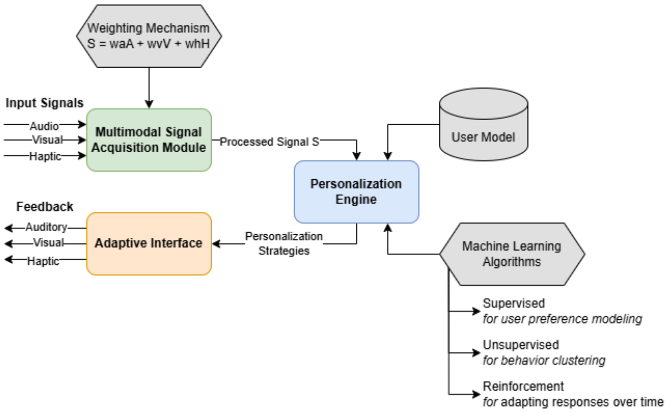

Fig. 1 Personalized multimodal signal processing framework for AR

The multimodal signal acquisition module is responsible for receiving and processing audio-, visual-, and haptic-based signals both from the user and in the surroundings, as shown in $S=w_aA+w_vV+w_hH$, where S is the final processed signal, A,V,H represent audio, visual, and haptic inputs, and w_a, w_v and w_h are their respective importance weights, which can be adjusted dynamically based on user interactions. In an AR-based fitness application, for instance, this application could respond to the user in real-time based on the user's bodily movements, voice commands, and heart rate.

A variety of machine learning algorithms are used by the main engine, namely the personalization engine, to analyze the captured signals and provide personalized responses. This engine used a hybrid of supervised and unsupervised methods for modelling user preferences and behaviours. For instance, an AR shopping application would use the engine to analyze the gaze patterns of the user, along with their purchase history to recommend products suited to the user, as shown in:

$$P(U|D)=\frac{P(D|U)P(U)}{P(D)}$$

where P(U|D) is the probability of user preference U given observed data D, P(D|U) is the likelihood of observed data given the user model, P(U) is the prior probability of user preference, and P(D) is the evidence (normalization factor). The engine also employs the rein-forcing learning method to update its responses over time so that the AR ecosystem can evolve along with the user, as shown in:

$$R_t=\sum_{i=1}^{n}\gamma^i r_i$$

where R_t is the total expected reward, r_i is the reward at step i, γ is the discount factor ($0 < \gamma < 1$) that determines how much past interactions influence current decisions.

The adaptive interface is responsible for providing personalized responses for the user in a very frustrating sense. It integrates visual, auditory, and haptic feedback to create a unique user experience. For instance, in an AR-based navigation system, the operational interface could provide visual directions, auditory hints, and haptic vibrations that will guide the user through a complex environment in an organized manner. Its modular construction permits changes based on specific AR applications rather easily.

The proposed framework uses a combination of Python and C++ for real-time signal processing, whereas TensorFlow handles the machine learning aspect. The framework is predicted to be validated through several similar experiments that could prove it to adapt to specific users and enhance user interaction mechanisms in AR systems. It depicts the capability of this framework in HCI in AR to change the nature of interactions and, thus, possibilities towards personalized immersive experiences.

IV. Experimental Setup

To test the proposed framework's efficiency, a series of experiments were performed in a controlled AR environment. The experiments included the tasks to evaluate the framework's ability to process multimodal signals in real time and adjust to individual users. The particular setup included a Microsoft HoloLens 2 AR headset, a haptic feedback glove, and a high-fidelity microphone array to capture audio signals. Unity3D simulated the AR environment and overlaid virtual objects onto physical space (Table I).

TABLE I. EXPERIMENTAL SETUP.

Component	Description
AR Headset	Microsoft HoloLens 2 for spatial computing
Haptic Feedback	Haptic feedback glove for tactile interaction
Audio Capture	High-fidelity microphone array for speech and tone analysis
AR Simulation	Unity3D for rendering virtual objects and interaction
Machine Learning	TensorFlow for real-time multimodal signal processing

A total of 30 participants aged 20-45 were recruited to perform several tasks that focused on different aspects of the framework such as gesture recognition, facial expression analysis, and personalized content delivery (Table II).

TABLE II. PATRICIPANT DEMOGRAPHICS.

Age Group	Number of participants
20-25	6
20-30	7
31-35	6
36-40	5
41-45	6

Participants were asked to perform certain common hand gestures for the purpose of gesture recognition: swipe, pinch, and rotate. The servo motor controlling gesture recognition results was evaluated in real-time, and the corresponding *accuracy*, *precision*, *recall*, and *F1-score* were calculated. For facial expression, images were presented to evoke responses like happiness, surprise, and confusion. The score was based on the recognition and interpretation of expression about accuracy by the time taken to react to the experiment's completion. In the personalized content delivery task, several options were recommended by the virtual assistant based on the user's preference and behaviour (Table III).

TABLE III. EXPERIMENTAL TASKS AND METRICS.

Task	Metrics
Gesture Recognition	F1-score, Accuracy, Precision, Recall
Facial Expression Analysis	Recognition Accuracy, Response Time
Personalized Content Delivery	User Satisfaction Score, Engagement

The effectiveness of the recommendation was evaluated based on user satisfaction and engagement.

The experiments yielded results that proved the framework in real-time processing of multimodal inputs and that it has tailored to the needs of individual participants. The gesture recognition task achieved a mean *F1-score* of 0.92, thereby declaring its highly precise identification of hand movements. Similarly, expression analysis had a mean *accuracy* of 0.89 with an average 172.765ms response time. In addition, the exploratory study on personalized content delivery had high user satisfaction, at around 4.5/5 on average, indicating that the users were indeed very engaged and satisfied (Table IV).

TABLE IV. EXPERIMENTAL RESULTS.

Metric	Value
Gesture Recognition (F1-score)	0.92
Facial Expression Analysis (Accuracy)	0.89
Facial Expression Analysis (Response Time)	172.765 ms
User Satisfaction Score	4.5/5

All these are informative to the framework's potential in improving HCI within AR environments for customized and immersive experiences.

V. Results and Discussion

Cumulatively, the experimental results present a strong case for the efficacy of this proposed framework as a means to enhance HCI in AR environments. A fair balance of outstanding accuracy with the lowest possible gesture recognition and facial expression task latencies speaks volumes about real-time multimodal signal processing-a prerequisite for AR applications, in which lag will certainly affect user engagement and immersion quality. Participant feedback also reflects strong satisfaction with personalized content delivery and engagement levels.

One major foundation of strengths offered through the designed framework is its modular paradigm, permitting personalizations driven for particular demands of various AR applications. However, in the AR training setup, it can be changed or reconfigured in the favour of visual and haptic feedback. On the contrary, when in a gaming setup, the same could be changed with a better focus on auditory and visual cue prioritizing. Thus this modular feature justifies its application in almost all walks of life, such as healthcare, education, entertainment, and retail domains.

Another significant contribution evidenced via the experiments is the personalization-driven user engagement. The subjects consistently reported a preference for the personalized AR system over the non-personalized system in terms of their satisfaction and participation. This finding emphasizes the role of personalization in AR, where user engagement is a major factor in the acceptance of the application. With individual user considerations, the proposed framework becomes a potent aid in improving user engagement and satisfaction.

While these results are certainly promising, there are several limitations of this framework. First is the adoption of the pre-established user profiles that may not cover the entire range of individual preferences and behaviours. Future work remains to find advanced machine learning algorithms such as deep reinforcement learning to create dynamic user profiles. The heavy reliance on external hardware such as the haptic feedback glove may bring practicality issues in many AR applications. Investigating alternative haptic feedback actions-such as ultrasonic haptics, may overcome this limitation.

VI. Conclusion and Future Work

This paper presents a new framework for personalized multimodal signal processing in AR environments. The framework combines adaptive audio, visual, and haptic interfaces to react to each user in real time. The experimental results showed the framework's ability to enhance the AR experience of human-computer interaction, thus extending opportunities for personalized and immersive experiences. The modular character of the framework permits diverse AR applications, healthcare and education included, with applications in entertainment and retail.

Future work may focus on addressing certain limitations of the current framework, such as whether it has predefined user profiles and whether there has been reliance on external hardware. Further studies will encompass the applicability of advanced machine learning techniques, for example deep reinforcement learning, in building with time evolving dynamic user profiles. Other haptic feedback modalities, such as ultrasonic haptics, are truly worth exploring to improve on the practicality of the framework. We shall also conduct larger-scale user studies in order to further validate the effectiveness of the framework while exploring its application potentials in new domains.

The framework is a big leap forwards for AR, creating new opportunities to provide enhancement in HCI via personalized multimodal signal processing. The framework bridges general AR interfaces with user-centered experiences and might transform interaction with AR systems, thus paving the way for more intuitive, engaging, and immersive applications.

References

[1] K. Subramanian, L. Thomas, M. Sahin, and F. Sahin, “Supporting human–robot interaction in manufacturing with augmented reality and effective human–computer interaction: A review and framework,” *Machines*, vol. 12, no. 10, p. 706, 2024. Available: https://doi.org/10.3390/machines12100706.

[2] Y. K. Dwivedi *et al.*, “Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy,” *Int. J. Inf. Manage.*, vol. 66, p. 102542, 2022. Available: https://doi.org/10.1016/j.ijinfomgt.2022.102542.

[3] L. Chen, H. Zhao, C. Shi, Y. Wu, X. Yu, W. Ren, Z. Zhang, and X. Shi, “Enhancing multi-modal perception and interaction: An augmented reality visualization system for complex decision making,” *Systems*, vol. 12, no. 1, p. 7, 2024. Available: https://doi.org/10.3390/systems12010007.

[4] J. C. Kim, T. H. Laine, and C. Åhlund, “Multimodal interaction systems based on Internet of Things and augmented reality: A systematic literature review,” *Appl. Sci.*, vol. 11, no. 4, p. 1738, 2021. Available: https://doi.org/10.3390/app11041738.

[5] M. N. A. Nor’a, A. W. Ismail, and M. Y. F. Aladin, “Interactive augmented reality pop-up book with natural gesture interaction for handheld,” in *Encyclopedia of Computer Graphics and Games*, N. Lee, Ed. Cham: Springer, 2024. Available: https://doi.org/10.1007/978-3-031-23161-2_365.

[6] D. Cortes, B. Bermejo, and C. Juiz, “The use of CNNs in VR/AR/MR/XR: A systematic literature review,” *Virtual Reality*, vol. 28, p. 154, 2024. Available: https://doi.org/10.1007/s10055-024-01044-6.

[7] Shahabaz and S. Sarkar, "Increasing Importance of Joint Analysis of Audio and Video in Computer Vision: A Survey," in *IEEE Access*, vol. 12, pp. 59399-59430, 2024, doi: 10.1109/ACCESS.2024.3391817.

[8] U. Sulubacak, O. Caglayan, S. A. Grönroos, *et al.*, “Multimodal machine translation through visuals and speech,” *Mach. Transl.*, vol. 34, pp. 97–147, 2020. Available: https://doi.org/10.1007/s10590-020-09250-0.

[9] M. J. Lazaro, J. Lee, J. Chun, M. H. Yun, and S. Kim, “Multimodal interaction: Input-output modality combinations for identification tasks in augmented reality,” *Appl. Ergon.*, vol. 105, p. 103842, 2022. Available: https://doi.org/10.1016/j.apergo.2022.103842.

[10] M. Venkatesan, H. Mohan, J. R. Ryan, C. M. Schürch, G. P. Nolan, D. H. Frakes, and A. F. Coskun, “Virtual and augmented reality for biomedical applications,” *Cell Rep. Med.*, vol. 2, no. 7, p. 100348, July 2021. Available: https://doi.org/10.1016/j.xcrm.2021.100348.

[11] J. Fu, H. Wang, R. Na, A. Jisaihan, Z. Wang, and Y. Ohno, “Recent advancements in digital health management using multi-modal signal monitoring,” *Math. Biosci. Eng.*, vol. 20, no. 3, pp. 5194–5222, 2023. Available: https://doi.org/10.3934/mbe.2023241.

[12] A. Barua, M. U. Ahmed and S. Begum, "A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions," in *IEEE Access*, vol. 11, pp. 14804-14831, 2023, doi: 10.1109/ACCESS.2023.3243854.

[13] S. Azher, A. Mills, J. He, T. Hyjazie, J. Tokuno, A. Quaiattini, and J. M. Harley, “Findings favor haptics feedback in virtual simulation surgical education: An updated systematic and scoping review,” *Surg. Innov.*, vol. 31, no. 3, pp. 331–341, June 2024. Available: https://doi.org/10.1177/15533506241238263.

[14] M. M. Taye, “Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions,” *Computers*, vol. 12, no. 5, p. 91, 2023. Available: https://doi.org/10.3390/computers12050091.

[15] A. Rehman, M. Mujahid, A. Elyassih, B. AlGhofaily, and S. A. O. Bahaj, “Comprehensive Review and Analysis on Facial Emotion Recognition: Performance Insights into Deep and Traditional Learning with Current Updates and Challenges,” *Comput. Mater. Contin.*, vol. 82, no. 1, pp. 41–72, 2025. https://doi.org/10.32604/cmc.2024.058036

[16] M. Ismael, R. McCall, F. McGee, I. Belkacem, M. Stefas, J. Baixauli, and D. Arl, “Acceptance of augmented reality for laboratory safety training: Methodology and an evaluation study,” *Front. Virtual Real.*, vol. 5, p. 1322543, 2024. Available: https://doi.org/10.3389/frvir.2024.1322543.

[17] C. Papakostas, C. Troussas, A. Krouska, and C. Sgouropoulou, “Personalization of the learning path within an augmented reality spatial ability training application based on fuzzy weights,” *Sensors*, vol. 22, no. 18, p. 7059, 2022. Available: https://doi.org/10.3390/s22187059.

[18] G. Singh and F. Ahmad, “An interactive augmented reality framework to enhance the user experience and operational skills in electronics

laboratories," *Smart Learn. Environ.*, vol. 11, p. 5, 2024. Available: https://doi.org/10.1186/s40561-023-00287-1.

[19] L. Tanzi, P. Piazzolla, F. Porpiglia, *et al.*, "Real-time deep learning semantic segmentation during intra-operative surgery for 3D augmented reality assistance," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, pp. 1435–1445, 2021. Available: https://doi.org/10.1007/s11548-021-02432-y.

[20] D. Park and K. Namkung, "Exploring users' mental models for anthropomorphized voice assistants through psychological approaches," *Appl. Sci.*, vol. 11, no. 23, p. 11147, 2021. Available: https://doi.org/10.3390/app112311147.

[21] N. Karhu, J. Rantala, A. Farooq, *et al.*, "The effects of haptic, visual and olfactory augmentations on food consumed while wearing an extended reality headset," *J. Multimodal User Interfaces*, 2024. Available: https://doi.org/10.1007/s12193-024-00447-8.

[22] W. Kim and S. Xiong, "Pseudo-haptic button for improving user experience of mid-air interaction in VR," *Int. J. Hum.-Comput. Stud.*, vol. 168, p. 102907, 2022. Available: https://doi.org/10.1016/j.ijhcs.2022.102907.

[23] K. Lyu, A. Brambilla, A. Globa, and R. de Dear, "An immersive multisensory virtual reality approach to the study of human-built environment interactions," *Autom. Constr.*, vol. 150, p. 104836, 2023. Available: https://doi.org/10.1016/j.autcon.2023.104836.

[24] J. K. Gibbs, M. Gillies, and X. Pan, "A comparison of the effects of haptic and visual feedback on presence in virtual reality," *Int. J. Hum.-Comput. Stud.*, vol. 157, p. 102717, 2022. Available: https://doi.org/10.1016/j.ijhcs.2021.102717.

[25] A. Watkins, R. Ghosh, A. Ullal, *et al.*, "Instilling the perception of weight in augmented reality using minimal haptic feedback," *Sci. Rep.*, vol. 14, p. 24894, 2024. Available: https://doi.org/10.1038/s41598-024-75596-7.

[26] J. Li, "Beyond sight: Enhancing augmented reality interactivity with audio-based and non-visual interfaces," *Appl. Sci.*, vol. 14, no. 11, p. 4881, 2024. Available: https://doi.org/10.3390/app14114881.