

Received October 7, 2021, accepted October 23, 2021, date of publication November 1, 2021, date of current version November 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3124607

Multi-Sensorial Human Perceptual Experience Model Identifier for Haptics Virtual Reality Services in Tactful Networking

BENEDETTA PICANO , (Member, IEEE)

Department of Information Engineering, University of Florence, 50139 Firenze, Italy

e-mail: benedetta.picano@unifi.it

ABSTRACT The tactful networking paradigm is expected to play a crucial role in the next generation networks. Accordingly, adaptive human-aware environments, sensitive to the daily human behavior and individual traits have to be provided, in order to offer a fully immersive and customized experience to users. On the basis of data collected by actual cognitive experiments, this paper proposes a learning framework to discover the multi-sensory human perceptual experience. The paper applies the mixture density network to identify the perception model considering different senses, and then the multi-sensory integration is performed, accordingly to the actual neuro-cognitive model. Furthermore, a supervised learning module has been used to cluster the users on the basis of the human perception identification strategy previously designed, assuming a multimodal structure for the cognitive brain activity. Finally, a practical contextualization is presented, in relation to the haptics virtual reality services. What emerges from the results is the effectiveness of the tactful approach, i.e., brain-aware, involving the proposed framework, which is validated in comparison to the more conventional brain-agnostic scheme. In fact, the system performance, expressed in terms of reliability in guaranteeing the service exploitation before a target deadline based on the integrated perception, reaches remarkable improvements applying the brain-aware strategy, which exploits the human perception knowledge.


INDEX TERMS Human-in-the-loop, quality-of-experience, tactful networking, supervised learning.

I. INTRODUCTION

One of the most salient features of the next generation networks is expected to be the need to manage human-aware applications, for example ultimate virtual reality services, or more in general those involving the extended reality (XR), or haptic communications. In fact, the new era of wireless networks will be characterized by architectures, communication models and technological solutions able to guarantee services based on daily humans behavior, including the psychological and cognitive aspects of the human brain, as well as the everyday humans habits, in order to meet the users expectations more naturally [1]. The recently emerged tactful networking paradigm has marked a divide between the user-centric applications, typical of the previous network generations, and the new era human-aware perspective, denoting a novel interdisciplinary area referred to the human behavior analysis. The user-centric approach is interested in whether and how

the user participation may help to solve the challenges of the networks, in order to guarantee the users' target requirements satisfaction. Differently, with the advent of the tactful networking, which boosts the human-in-the-loop vision, the attention has been posed to the extraction and analysis of the human-traits, expressed in terms of human perception, capability in observation, interpretation, and reaction of the humans to the surrounding environment. Looking forward, by involving the human subjective sphere of the individuals within the new era networks and applications design, such as the human personality, the routines, the brain cognitive limitations, and so on, the quality-of-experience (QoE) may be empowered, offering the chance to realize tactful network ecosystems, i.e., environments adaptive to the human context, sensitive to human behavior and interests, and able to support real-time and interactive virtual environments soliciting the users' five senses [1].

Within this context, the exploitation of advanced machine learning techniques to deeply investigate the cognitive aspects of the human brain and the behavioral individual

The associate editor coordinating the review of this manuscript and approving it for publication was Javed Iqbal .

traits, is gaining momentum, expecting to play a crucial role in next generation networking [1]. In reference to this, some pioneer multidisciplinary studies have been proposed, for example [2] and [3], in which the users' delay perception has been analyzed and substantial efforts have been made to design a framework able to catch the human perception dynamics in relation to a wide range of individual features, such as age, personality, etc., representing remarkable literature improvements. What emerges in papers [2] and [3] is that there exist individual human brain limitations in perceiving quality-of-service (QoS) improvements above a certain threshold, leading to biases in how the human brain translates the QoS in the actual QoE level, perceiving a QoE lower than the expected one [2]. In fact, the recent multi-disciplinary literature has started highlighting the actual cognitive limitations of the human brain in perceiving different QoS in video streaming transmission, in reference to both the rate and the delay metrics [2], [4], [5], confirming that human users cannot perceive QoS gains due to the intrinsic cognitive constraints.

Nevertheless, the challenges concerning the design of models and frameworks to realistically fit the actual capability of the human behavioral perception, in order to incorporate it in the practical realization of XR environments soliciting several senses simultaneously, are still numerous. For example, due to the intrinsic complexity of the data collection and the multifacet nature of the topic, literature lacks in models to identify human perceptions different from that related to the visual cortex cognitive awareness.

This paper proposes the formulation of a statistical learning framework to model the human perception of the users considering different senses. More in depth, the contributions of the paper in comparison to the existing literature can be summarized as follows

- A multidisciplinary machine learning based framework, which aims at providing a model to identify the multi-sensorial delay perceptions of the human brain, considering actual data deriving from cognitive science trade journals experiments as [6]–[8], and credited neuro-cognitive perceptive brain models [9]–[13];
- The application of a Gaussian mixture based learning method to identify the human perception related to different sensory channels to cluster users in relation to the Gaussian mode which better fits the corresponding perception dynamic;
- The design of a framework to model the multi-sensorial human perception experience considering the integration of different sensory channels stimulated simultaneously, and the potential dominance of the visual perception on the haptic cue. The proposed framework has been contextualized to a tactful network arranged to handle haptics virtual reality (HVR) services, to show the advantages of performing human-perception based solutions, in reference to the system reliability in guaranteeing the service exploitation before a target deadline.

The rest of the paper is organized as follow. Section II presents the review of the prior works related to our study. Section III describes the proposed human perception identifier model. Then, the Section IV presents the contextualization of the proposed model to a HVR scenario and the corresponding performance analysis. Finally, the conclusions are drawn in Section V.

II. RELATED WORKS

Several works have been proposed involving the users' personal information, such as the behavioral patterns or the social interactions, into frameworks devoted to address the traditional problems of the modern networks. Examples are represented in papers [15], [19], and [2]. The user-centric vision was adopted in paper [15], where authors aim at optimizing the resource allocation procedure in wireless small cell networks, throughout matching theory, considering device-to-device communication. Moreover, authors in [15] apply matching theory considering as relevant feature the social aspects, providing a context-aware resource allocation framework based on the social interactions. Studies about the human perception have been proposed in [19], in which the authors design a spectrum sharing strategy on the basis of the human psychological behavior of the users involved in the network. Paper [2] presents a power minimization considering the limitations of the human brain in perceiving different QoS levels, applying the Lyapunov optimization.

The multi-sensory perception has been the focus of paper [20], in which a multi-sensory learning framework has been designed to convert text sentence into visual and auditory representations to support autistic students. The paper [21] conducts an experiment about the human perception, in reference to both the visual and proprioceptive stimuli. The experiment is focused on 8 subjects with trained proprioception and 8 subjects with visual training. The corresponding results highlight that the learning rate of the visual perception is remarkably higher than that related to the proprioception; although the statistical error results to be decreasing as the learning training increases, the error still results significant for proprioception and non-significant for visual channel. Furthermore, the error in the test phase results to be higher than in the training because during the test, the process involved is the multi-sensory integration. The authors in paper [22] address the problem of the human-computer interaction, aiming at providing the state-of-the-art in data representation involving more than one sensory channel. The paper has considered 154 examples of multisensory data representations, in order to propose a design space along three dimensions: use of modalities, representation intent and human–data relations. In the paper [23], a control system prototype based on Arduino has been designed for the heat and scent simulations, within the VR environments. The proposed system represents a multisensory VR simulator developed in Unity 3D to prevent hazards such as fire, smoke, and so on. The paper [24] exploits a multilayer perceptron architecture to classify unisensory stimuli, on the basis of

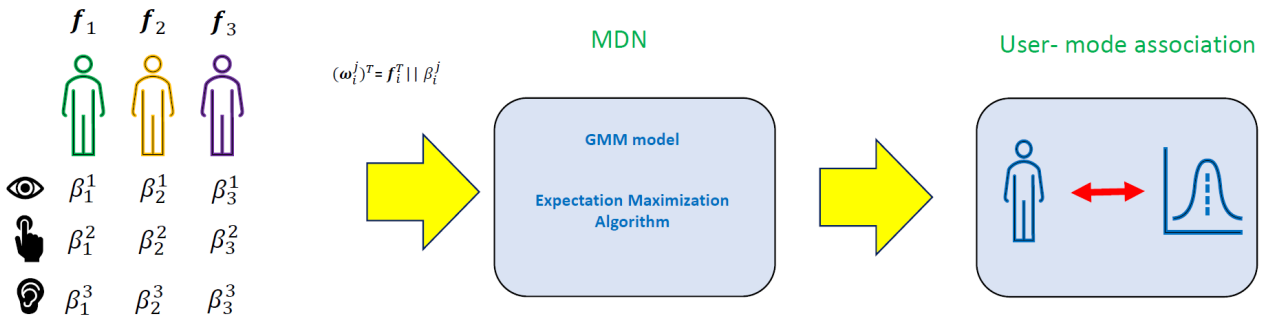


FIGURE 1. Uni-sensory most probable mode identification.

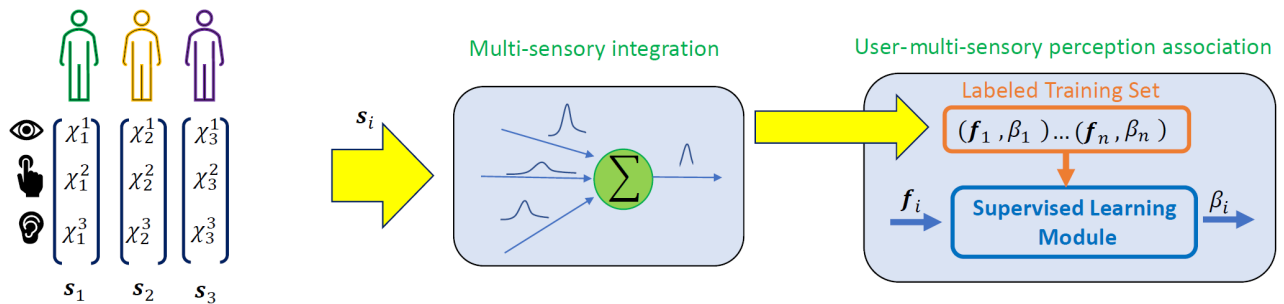


FIGURE 2. Multi-sensory human perception identification framework.

which, then, a generalized feature-integrating model is proposed to analyze the multisensory combination. Furthermore, the development of a multi-sensory augmented reality system for the cultural heritage has been performed in [25], aiming at evoking several stimuli with SensiMAR prototype. The proposed system is referred to offer a multi-sensory experience involving the visual reconstructions, soundscape of ancient times, and smell very common during the historical period considered.

This paper aims at pursuing a multi-disciplinary approach to the multi-sensory human perception, in order to provide a valuable integrated multi-sensory model to be exploited in tactful networking applications. Nevertheless many studies have been conducted in relation to human visual-haptics combination, this paper has the goal to merge the neurocognitive aspects of the human brain with the machine learning techniques in order to provide a comprehensive framework for human perceptual experience to be exploited in tactful networking. In this sense, for the best of author’s knowledge, there is not yet a similar work in the existing literature.

III. HUMAN PERCEPTUAL EXPERIENCE MODEL IDENTIFIER

As it is conspicuous from literature [2], [3], [6]–[8], the human cognitive perception does not result from only one distinctive trait of the individuals, rather it derives from a multitude of attributes, hereafter referred as features, such as gender, age, diseases, or the distance from a reference point.

In this paper the presence of diseases inhibiting one or more sensory channels is not taken into account, as well as the condition in which the sensory stimuli are not referred to the same scene. In order to design and support reactive multi-senses soliciting XR environments, the match between user features profiles and the corresponding cognitive perception has to be performed. By focusing first of all on only one sensory channel j , let $\mathcal{F} = \{1, \dots, n\}$ be the set of users, and let $\{f_1, \dots, f_n\}$ be the set of the features vectors corresponding to n users, respectively. Consequently, $f_i \in \mathbb{R}^m$ represents the features vector of the i -th user and, for each f_i , $i = 1, \dots, n$, F^j represents the features matrix, in which entry i is f_i^T . Then, the vector β^j is the user uni-sensory perception vector, whose element $\beta_i(t)^j$ expresses the perception of user i in reference to the j -th human sense. The features vectors can be collected from neurocognitive experiments focusing on each sense, such as those in [6]–[8]. In accordance with Fig. 1 and Fig. 2, the proposed learning framework consists of an unsupervised learning module implementing the mixture density network (MDN) estimation [26], and of a supervised learning solution to cluster the users on the basis of the features they exhibit. More in depth, the unsupervised learning module allows to match the users’ features vectors f_i with the most probable perception mode, supposing a user uni-sensory perceptual experience. Then, a strategy for the multi-sensory integration is proposed, on the basis of the uni-sensory perceptual values previously considered. In this way, a labeled dataset in which each features vector is classified on

the basis of the multi-sensory perception value is produced. Therefore, the labeled dataset is used to train a supervised learning module aiming at classifying the users on the basis of their multi-sensory perceptual experience value, in accordance with the corresponding features vector.

A. UNSUPERVISED UNI-SENSORY PERCEPTUAL EXPERIENCE IDENTIFICATION MODEL

The unsupervised learning block consists of the MDN module that adopts the Gaussian mixture model (GMM) and the Expectation Maximization algorithm (EM) to solve the unsupervised clustering problem [2], [27]. Accordingly to the statistical learning [27], the mixture model exploits the Bayes rule to give the degree of membership of an element to any cluster, hereafter referred as brain mode, by resorting to the conditional probabilities. More in depth, the MDN module allows to model the conditional probability distribution $p(\omega_i^j|f_i)$, where $\omega_i^j \in \mathbb{R}^{m+1}$ represents the vector including $\beta_i(t)^j$ and the corresponding features f_i . From the GMM model assumption follows that we have a normalized linear combination of K Gaussian distributions. Therefore, we obtain that [2], [27]

$$p(\omega_i^j) = \sum_{k=1}^K p(z) p(\omega_i^j|z) = \sum_{k=1}^K \pi_k N(\omega_i^j|\mu_k^j, C_k^j), \quad (1)$$

in which $N(\omega_i^j|\mu_k^j, C_k^j)$ represents the k -th Gaussian component density characterized by mean vector μ_k^j and covariance matrix C_k^j , considering the sensory channel j . Furthermore, by introducing the latent binary random vector z , in which only one component is set to one and the other are zero, we have that $p(z_k = 1) = \pi_k$, which expresses the Gaussian component activated. More in depth, under the assumption of a multi-modal cognitive human activity, z gives information about the Gaussian mode fitting the human brain. In order to cluster data on the basis of the most probable mode to which they belong to, the EM algorithm aims at finding π_k , μ_k^j and C_k^j to maximize the log likelihood function expressed by [27]

$$\mathcal{L} = \sum_i \ln \sum_{k=1}^K \pi_k N(\omega_i^j|\mu_k^j, C_k^j). \quad (2)$$

Due to the complexity in maximizing (2), the EM acts in accordance with the following steps [27]

- 1) Randomly initialize π_k , μ_k^j and C_k^j ;
- 2) For each mode, compute the responsibility [27]

$$r_k^j = \frac{\pi_k N(\omega_i^j|\mu_k^j, C_k^j)}{\sum_{l=1}^K \pi_l N(\omega_i^j|\mu_l^j, C_l^j)}; \quad (3)$$

- 3) On the basis of the current responsibility value, the parameters are updated;
- 4) Repeat steps 1)-3) until convergence.

Consequently, the output of the EM, i.e., the unsupervised learning, is represented by the GMM corresponding to the

dataset matrix M_j , where $M_j = [F^j|\beta^j]$. Therefore, in reference to the obtained GMM model, each data sample ω_i^j , is labeled with the most probable Gaussian component k computed as follows [26]

$$\chi_i^j = \operatorname{argmax} p(z_k = 1|\omega_i^j) = \operatorname{argmax} r_k^j, \quad (4)$$

in which

$$p(z_k = 1|\omega_i) = \frac{\pi_k N(\mu_k^j, C_k^j)}{\sum_{l=1}^K \pi_l N(\mu_l^j, C_l^j)}. \quad (5)$$

At this stage, before resorting to the supervised learning module, we need to integrate together the output vectors of the unsupervised learning, i.e., $y^j = [\chi_1^j, \dots, \chi_n^j]^T$, for each sense j .

B. MULTI-SENSORY INTEGRATION

As previously detailed, the unsupervised learning module returns, for each sense j , the output vector $y^j = [\chi_1^j, \dots, \chi_n^j]^T$, expressing the most probable brain mode for each user. Then, let \mathcal{K}_k^j be the k -th cluster, then \mathcal{K}_k^j includes all the users' having as most probable mode χ_k^j , in reference to the sensory channel j . Therefore, among the elements belonging to \mathcal{K}_k^j , the user i^* can be identified as follows

$$i_k^* = \min \mathcal{B}_k^j, \quad (6)$$

where the set $\mathcal{B}_k^j = \{i \in \mathcal{F}|f_i \text{ has mode } \mathcal{K}_k^j\}$. Similarly, the user belonging to the k -th cluster exhibiting the maximum perception value can be defined as

$$i'_k = \max \mathcal{B}_k^j. \quad (7)$$

What it is intuitive to understand, the perception level of the users belonging to the k -th cluster, considering the sensory system j , is surely within the range $[\beta_{i'_k}^j, \beta_{i_k^*}^j]$. Due to the fact that lower is the delay perceived on a sensory channel, greater is the sensibility of the users in discerning the fluctuations occurred on that sense, it is reasonable to consider the worst case value $\beta_{i_k^*}^j$ as delay perception threshold for the considered k -th cluster of users when only one sensory channel is involved. In fact, the clustering method previously discussed may be useful to group together users with a common delay perception for a given sensory system. Nevertheless, whether and how the users' perception varies under the simultaneous stimulation of different sensory channels is still an open issues object of debate. Some studies have highlighted that sending stimuli contemporaneously on different sensory systems may originate interference on cognitive perception or even empower the human overall perception [9]–[13].

Although in many studies the advantages of the multi-sensory integration is emerged, other papers have highlighted that the brain, integrating information stemmed by some different channels, may lose the individual sensory system perception, leading to the metameric condition. Accordingly,

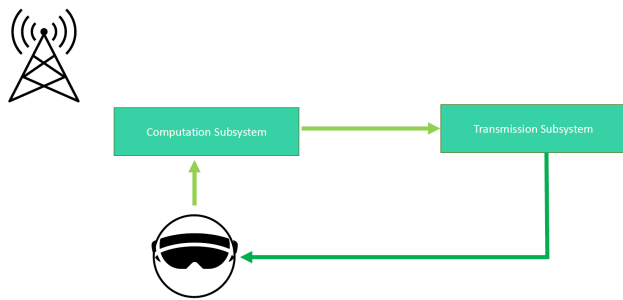


FIGURE 3. System scenario.

there may be different physical stimuli that lead to exactly the same perceptual experience, indiscriminable from one another [10].

Within this context, in order to model the multi-sensory human perception under simultaneous stimulation of different sensory channels, we can define, for each user i , the multi-sensory vector $s_i \in \mathbb{R}^J$, where J represents the number of sensory channels stimulated simultaneously. Furthermore, the entries in s_i represent the cluster to which they belong to, derived from the module previously described. Then, numerous neuro-cognitive studies have highlighted that the integration of the sensory cues are interpreted by the brain as a weighted sum of individuals perceptions. Consequently, a prior estimation of the multi-sensory perception β_i may be represented by the following expression

$$\beta_i = \sum_{j=1}^J w_j \mu_i^j, \tag{8}$$

where the weight w_j has to be properly defined, for example as the normalized reciprocal variance:

$$w_j = \frac{1/\sigma_j^2}{\sum_{j=1}^J 1/\sigma_j^2}. \tag{9}$$

Consequently, the integrated perception β_i results to be given by the sum of the uni-sensory perception, weighted with the reliability corresponding to each perception estimation, since the reciprocal of the variance is a measure of the reliability of such an estimation.

Nevertheless, other experimental studies as [11], [28], [29] have brought to light the existence of a relationship of *dominance* and *recessive* among the different sensory channels. In particular, the visual perception has been recognized as dominant when it is integrated with the haptic channel and the variance associated with visual estimation is lower than that associated with haptic estimation [11]. In this case, the weight associated to the visual system w_V is set as

$$w_V = w_H \frac{\sigma_H^2}{\sigma_V^2}, \tag{10}$$

in which w_H represents the weight given to the haptic cue and σ_H^2 and σ_V^2 are the variance of the haptic and visual modalities, respectively.

C. SUPERVISED USER-MULTISENSORY PERCEPTION ASSOCIATION

Once the multi-sensory integration is provided, we dispose of pair $\{f_i, \beta_i\}$, in which β_i depends on the elements y_i^j , for each j , of the output vectors y^j , i.e., s_i . Thus, the pair $\{f_i, \beta_i\}$ is used to train the supervised learning module which realizes a classifier that, for each human features vector f_i as input, returns the corresponding integrated perception value β_i . Let $\beta = [\beta_1, \dots, \beta_n]^T$ be the vector of the integrated perception values. Given the matrix F^j and β , the supervised learning module builds a model g such that $\beta_i = g(f_i)$, in which

$$g = \operatorname{argmin}_{\hat{g}} \sum_{i=1}^n \zeta(\beta_i, \hat{g}(f_i)), \tag{11}$$

and $\zeta(\cdot)$ is a 0–1 loss function [2], [27], and g is approximated exploiting the points (f_i, β_i) . In addition, to prevent overfitting, the Elbow method has been applied [2], [27]. Figure 4 summarized the proposed whole framework, posing emphasis on the relationship existing between the different modules of the framework.

Therefore, the proposed framework, represented in Figure 2, acts as follows

- 1) For each features vector f_i , and each sensory channel j stimulated, the GMM model based on ω_j^i is applied;
- 2) For each user i and sensory channel j , χ_i^j is computed;
- 3) For each most probable mode χ_i^j , the variance σ_j^2 is calculated;
- 4) For each user i , the multi-sensory perception e is computed by resorting to the brain multi-sensory integration expressed by (8);
- 5) The model described in (8) and (9) is applied when $\sigma_V > \sigma_H$ to find β_i ; otherwise the multi-sensory perception β_i is computed applying (8), (9), and (10);
- 6) At this point, a supervised learning module is applied to cluster the users accordingly to their β_i , on the basis of the features vector f_i . More in depth, in reference to Fig. 2, this module aims at associating to the feature vector f_i of each user, the corresponding effective delay perception β_i . In fact, steps 1) – 5) produce a labeled dataset which can be used for training the supervised learning module, making possible the association between users' features vectors f_i and the corresponding perception values β_i .

In the following section a case of study related to the VR haptics services is presented, exploiting the multi-sensory human perceptual experience model identifier proposed.

IV. BRAIN-AWARE END-TO-END DELAY ANALYSIS FOR HAPTICS VR SERVICES

The problem presented here is contextualized to the HVR services, which is one of the most disruptive applications among those expected in the next generation networks. It is important to highlight that the analysis proposed in the following has the only purpose of providing a contextualization to the framework previously designed. Therefore, the markovian analysis

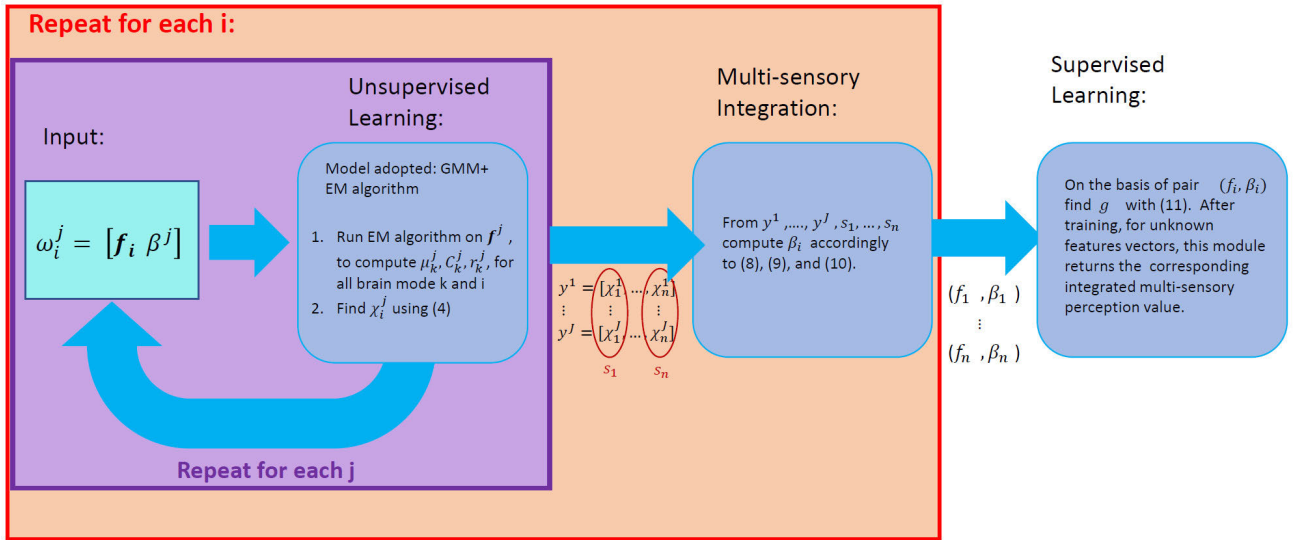


FIGURE 4. Proposed framework.

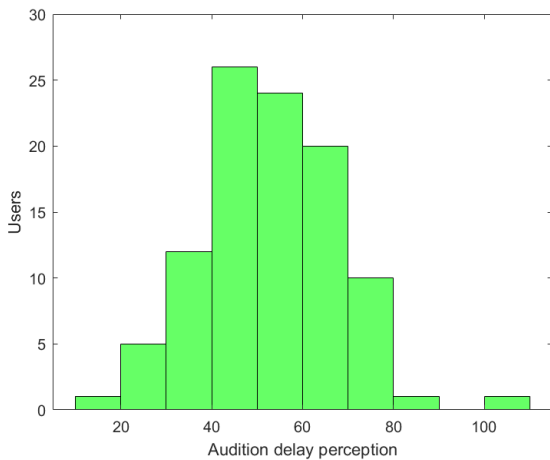


FIGURE 5. Distribution of audition perception considering 100 users.

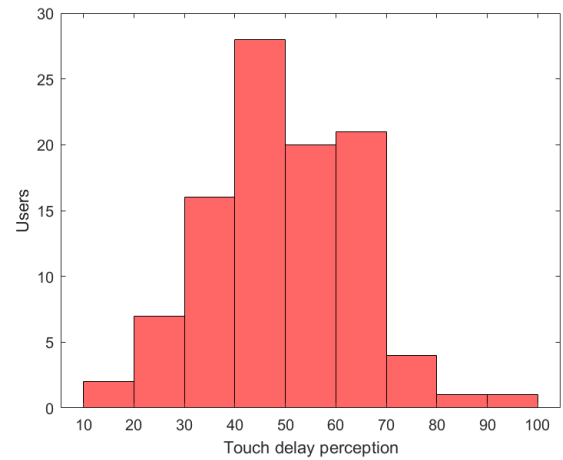


FIGURE 6. Distribution of touch perception considering 100 users.

TABLE 1. Simulation parameters.

| Parameter | Value |
|----------------------------------|------------------|
| \bar{x}_1 | 5ms |
| \bar{x}_2 | 0.75ms |
| Q | [1, 5] |
| \hat{k} | $0.0016m^{-1}$ |
| Bandwidth | 13GHz |
| Molecular absorption coefficient | $0.0016m^{-1}$ |
| Frequency | 1THz |
| Network dimension | 20×20 m |

presented represents a simplification of the problem in its complete form, and it has been given in order to provide an insight about the practical application of the multi-sensory perception model developed. Nevertheless, the proposed contextualization can be extended to more complex and detailed analysis.

Typically, HRV services solicit the vision, the audition, and the touch. In order to provide a useful application example of the proposed framework, we refer to the scenario considered in [34] and [35], whose main parameters are reported in Table 1. In reference to Fig. 3, a set of Q HVR users has been considered. Each user injects in the network a Poisson traffic flow with mean rate λ_q packets/ms, $q = 1, \dots, Q$. The communication link is provided by a SBS, expected to operate at terahertz frequencies, which represent one of the most promising option to guarantee communications characterized by simultaneously high rate, high reliability, and low latency for immersive VR experiences. After receiving computation, the packets are sent back to the user, throughout the transmission subsystem. The focus here is the evaluation of the delay due to the computing and transmission subsystem, i.e., the end-to-end delay, in reference to the delay

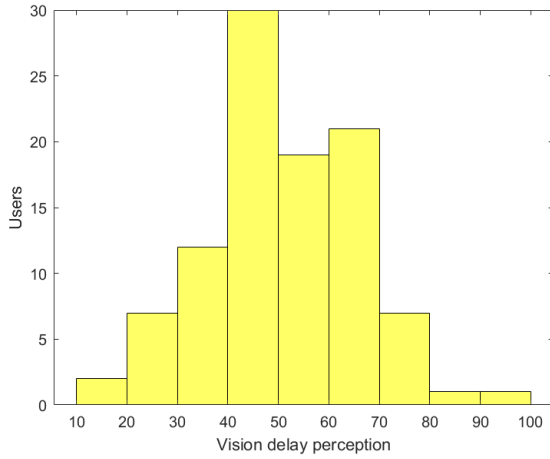


FIGURE 7. Distribution of vision perception considering 100 users.

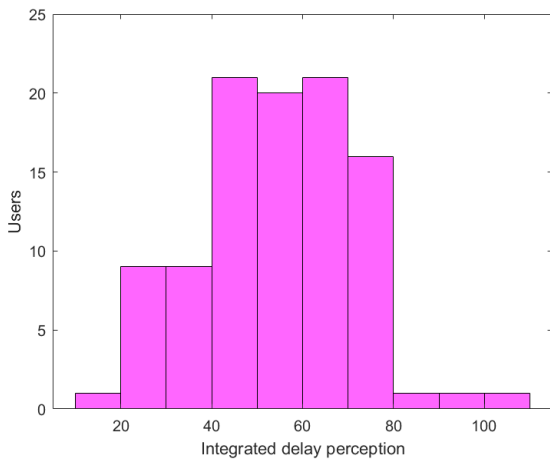


FIGURE 8. Distribution of integrated perception β_i considering 100 users.

target deadline \hat{k} . Both the computation and the transmission subsystems service times have been assumed as exponentially distributed, with mean service time $\bar{x}_1 = 5$ ms and $\bar{x}_2 = 0.75$ ms, in accordance with the reference scenario analyzed in [34], [36] and [35]. Hereafter the reliability of the system refers to the probability of guaranteeing the target deadline \hat{k} . Therefore, the HVR packets arrive at the computational subsystem with a Poisson process with mean rate $\lambda = Q\lambda_q$. The parameter of the exponential distribution modeling the service time of the computational subsystem is $\mu_1 = 1/\bar{x}_1$. Consequently, the corresponding processor load ρ is given by $\rho = \bar{x}_1\lambda_q$. The computation subsystem results to be a M/M/1 system. Due to the Burke's theorem [30], we have that the arrival process at the transmission subsystem results to be a Poisson process with mean rate λ , independent from that at the computational subsystem. Then, the exponential distribution modeling the service time of the transmission subsystem is $\mu_2 = 1/\bar{x}_2$.

Consequently, the overall e2e delay needed to accomplish a HVR request service, is given by the sum of two independent

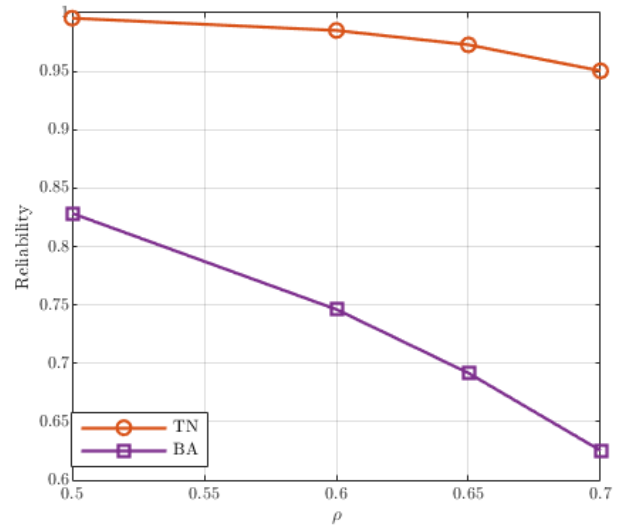


FIGURE 9. Reliability as a function of ρ .

time contributions: the time needed to complete service at the computation subsystem (t_1), and the time spent at the transmission subsystem (t_2) to send back to the HVR user the outcome of the service requested.

Let $\psi_1(t)$ and $\psi_2(t)$ be the pdf of the random variable t_1 and t_2 , respectively. The pdf of the overall e2e delay results to be given by the convolution

$$\pi(t) = \psi_1(t) * \psi_2(t).$$

In accordance with the standard queueing theory [30] we have

$$\begin{aligned} \psi_1(t) &= (\mu_1 - \lambda)e^{-(\mu_1 - \lambda)t} = \phi_1 e^{-\phi_1 t}, \\ \psi_2(t) &= (\mu_2 - \lambda)e^{-(\mu_2 - \lambda)t} = \phi_2 e^{-\phi_2 t}, \end{aligned} \quad (12)$$

where $\phi_1 = \mu_1 - \lambda$, and $\phi_2 = \mu_2 - \lambda$. After some algebraic manipulations we have that the reliability metric \mathcal{R} for the tandem model considered is

$$\mathcal{R} = \int_0^{\hat{k}} \pi(t)dt. \quad (13)$$

The uni-sensory model perception has been realized by considering the experimental data collected and merged by neuro-cognitive paper journals as [13], [31], [32], in reference to each sense involved in the HVR services. Consequently, the uni-sensory perception histogram corresponding to each sense involved in this application is reported in Fig. 5, Fig. 6, and Fig. 7, considering 100 users. Differently, Fig. 8 represents the histogram of the delay perception, considering 100 users.

On the basis of the integrated perception model obtained in Fig. 8, the value of \hat{k} may be set in accordance with the multi-sensory perception values estimated for the users experience and reported in that figure. It is important to highlight that the target deadline \hat{k} has been set equal to the statistical average of the perception values distribution

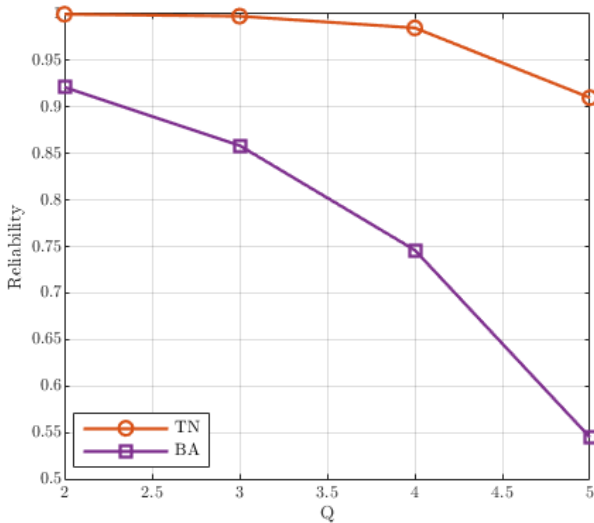


FIGURE 10. Reliability as a function of Q .

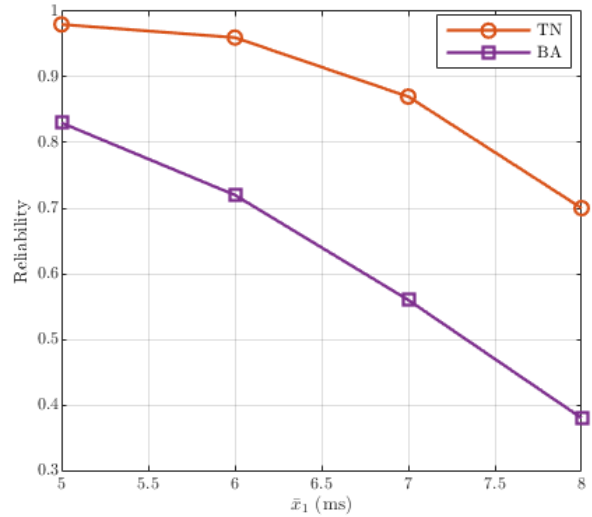


FIGURE 12. Reliability as a function of \bar{x}_1 .

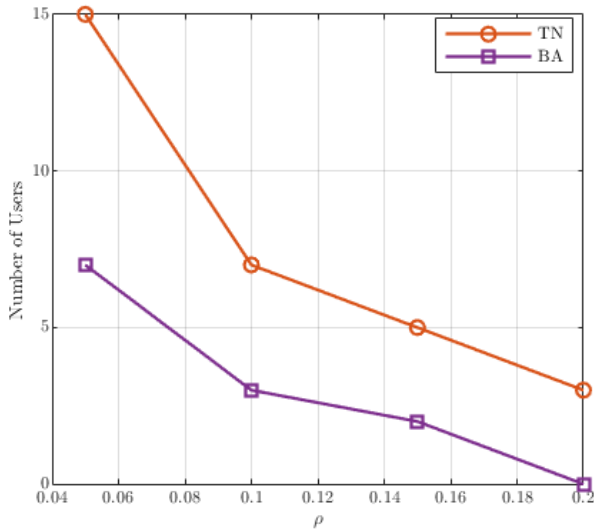


FIGURE 11. Reliability as a function of ρ .

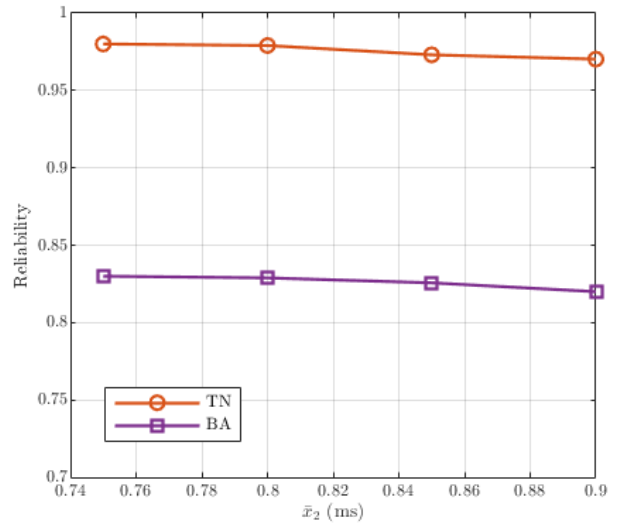


FIGURE 13. Reliability as a function of \bar{x}_2 .

illustrated in Fig. 8. In such a way, the target deadline results to be less stringent in comparison to the deadline imposed by the strict QoS requirements which, typically, is set to 20 as in standard literature [33], [34], [36] and [35]. In fact, the delay perception gives the opportunity to exploit more effectively the network resources and to offer simultaneously an immersive and fully satisfying XR experience to the users. Aiming at highlighting the benefits deriving from the tactful networking approach, considering the values presented in Fig. 8, Fig. 9 shows the system reliability as a function of the processor load ρ , considering $Q = 1$. The results are obtained by pursuing both the tactful networking (TN) and the more conventional brain-agnostic (BA) scheme, which considers $\hat{k} = 20$ ms. Differently, the TN approach sets the target deadline \hat{k} equal to the statistical average of the

obtained numerical values according to the perception values distribution illustrated in Fig. 8. As it is clearly evident from the Fig. 9, the TN strategy guarantees reliability values higher than those reached by applying the BA scheme. Such a trend is also confirmed by Fig. 10, in which the reliability is expressed as a function of the number of users Q . In this case, the arrival rate λ_q has been set the same for all the Q users and it is equal to 0.1 packets/ms. Also in this case, the results confirm the validity and the advantages of adopting a TN strategy, in order to offer more reliable HVR services and to handle a greater number of users. Fig. 11 considering a target reliability equal to 0.9, represents the number of users manageable by system without lowering reliability below the fixed target. As it is evident from the Fig. 11, also in this case, the improvements reached by applying the TN strategy confirms the validity of the human-aware approach. In addition,

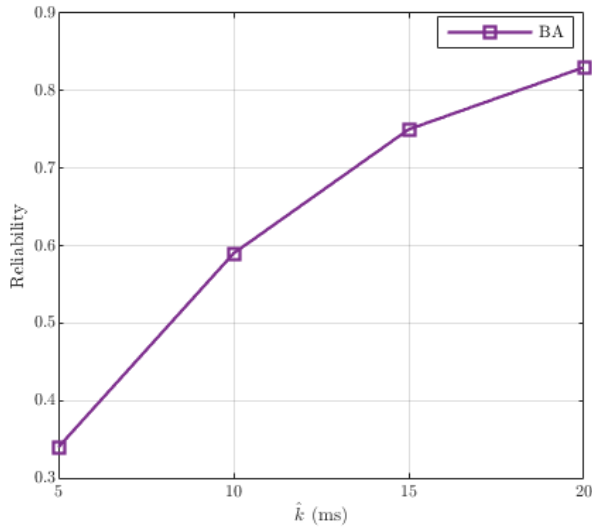


FIGURE 14. BA Reliability as a function of \hat{k} .

Fig. 12 and Fig. 13 express the system reliability as a function of the mean service time of the first (\bar{x}_1) and the second (\bar{x}_2) subsystem, respectively. From Fig. 12 and Fig. 13 is evident that the computation subsystem represents the bottleneck of the considered tandem network. In fact, the system reliability trend is steeper when \bar{x}_1 increases, in comparison to the reliability obtained by increasing the value of \bar{x}_2 . Finally, Fig. 14 shows the system reliability behavior for different values of \hat{k} , considering the BA approach. As it is straightforward to note, the reliability grows when we have less stringent deadline values.

V. CONCLUSION

This paper has proposed a learning based framework to model the human perception considering the five sensory systems, assuming the cognitive human brain activity as having a multimodal structure. The mixture density network has been applied to identify the conditional probability distribution of having, for each sense, a perception level, assuming some individual traits. Furthermore, a multi-sensory perception model has been assumed, in accordance to the neurocognitive literature, to perform sensory integration. Then, a supervised learning module has been used to cluster the users on the basis of their integrated perceptive sensitivity. Finally, the contextualization to the HVR services, in order to exhibit the advantages resulting from the adoption of a brain-aware decision making scheme based on the proposed framework, instead of a more common brain-agnostic approach, has been performed. The case of study presented aims at highlighting the performance improvements considering the system reliability in guaranteeing a perceptual target deadline.

REFERENCES

[1] R. L. Costa, A. C. Viana, A. Ziviani, and L. N. Sampaio, "Tactful networking: Humans in the communication loop," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 1, pp. 92–107, Feb. 2021.

[2] A. T. Z. Kasgari, W. Saad, and M. Debbah, "Human-in-the-loop wireless communications: Machine learning and brain-aware resource management," *CoRR*, vol. abs/1804.00209, pp. 1–33, Mar. 2018.

[3] Y. Yang, L. T. Park, N. B. Mandayam, I. Seskar, A. L. Glass, and N. Sinha, "Prospect pricing in cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 1, pp. 56–70, Mar. 2015.

[4] T. Zhao, Q. Liu, and C. W. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 285–302, 1st Quart., 2017.

[5] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1126–1165, 2nd Quart., 2015.

[6] J. Mojet, "Taste perception with age: Generic or specific losses in supra-threshold intensities of five taste qualities?" *Chem. Senses*, vol. 28, no. 5, pp. 397–413, Jun. 2003.

[7] R. Rupini and R. Nandagopal, "A study on the influence of senses and the effectiveness of sensory branding," *J. Psychiatry*, vol. 18, no. 2, p. 236, 2015.

[8] S. Kime, F. Galluppi, X. Lagorce, R. B. Benosman, and J. Lorenceau, "Psychophysical assessment of perceptual performance with varying display frame rates," *J. Display Technol.*, vol. 12, no. 11, pp. 1372–1382, Nov. 2016.

[9] H. B. Helbig and M. O. Ernst, "Optimal integration of shape information from vision and touch," *Exp. Brain Res.*, vol. 179, no. 4, pp. 595–606, Jun. 2007.

[10] M. Ernst, "A Bayesian view on multimodal cue integration," in *Human Body Perception From the Inside Out*, G. Knoblich, M. Grosjean, I. Thornton, and M. Shiffrar, Eds. Oxford, U.K.: Oxford Univ. Press, 2006, pp. 105–131.

[11] M. Ernst and M. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, pp. 429–433, Feb. 2002.

[12] M. O. Ernst and H. H. Bühlhoff, "Merging the senses into a robust percept," *Trends Cogn. Sci.*, vol. 8, no. 4, pp. 162–169, 2004.

[13] J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy, "Combining sensory information: Mandatory fusion within, but not between, senses," *Science*, vol. 298, no. 5598, pp. 1627–1630, Nov. 2002.

[14] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *CoRR*, vol. abs/1902.04885, pp. 1–19, Feb. 2019.

[15] J. Zhao, Y. Liu, K. K. Chai, M. Elkashtan, and Y. Chen, "Matching with peer effects for context-aware resource allocation in D2D communications," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 837–840, Apr. 2017.

[16] M. Alam, D. Yang, K. Huq, F. Saghezchi, S. Mumtaz, and J. Rodriguez, "Towards 5G: Context aware resource allocation for energy saving," *J. Signal Process. Syst.*, vol. 83, no. 2, pp. 279–291, May 2016.

[17] Y. Lin, R. Zhang, C. Li, L. Yang, and L. Hanzo, "Graph-based joint user-centric overlapped clustering and resource allocation in ultradense networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4440–4453, May 2018.

[18] M. Zalghout, S. Abdul-Nabi, A. Khalil, M. Helard, and M. Crussiere, "Optimizing context-aware resource and network assignment in heterogeneous wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[19] Y. Yang, L. T. Park, N. B. Mandayam, I. Seskar, A. Glass, and N. Sinha, "Prospect pricing in cognitive radio networks," *CoRR*, vol. abs/1510.00781, pp. 1–39, Dec. 2015.

[20] S. L. De Silva and N. Dissanayaka, "Multisensory learning approach to create a sentence learning platform for students with autism spectrum disorder," in *Proc. 3rd Int. Conf. Conver. Technol. (I2CT)*, Apr. 2018, pp. 1–5.

[21] M. Saidi, A. A. Lari, and F. Towhidkhan, "Comparison of visual and proprioceptive training on multisensory perception using a new designed setup," in *Proc. 21st Iranian Conf. Electr. Eng. (ICEE)*, May 2013, pp. 1–4.

[22] T. Hogan and E. Hornecker, "Towards a design space for multisensory data representation," *Interacting Comput.*, vol. 29, no. 2, pp. 147–167, May 2016.

[23] M. Tahsiri, G. Lawson, C. Abdullah, and T. Roper, "A multisensory virtual environment for OSH training," in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Mar. 2018, pp. 699–700.

[24] R. Amerineni, L. Gupta, and R. S. Gupta, "Classification models inspired by multisensory integration," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 255–258.

[25] A. Marto, M. Melo, A. Gonçalves, and M. Bessa, "Development and evaluation of an outdoor multisensory ar system for cultural heritage," *IEEE Access*, vol. 9, pp. 16419–16434, 2021.

- [26] J. Li and A. Barron, "Mixture density estimation," in *Advances in Neural Information Processing Systems*, vol. 12, S. Solla, T. Leen, and K. Müller, Eds. Cambridge, MA, USA: MIT Press, 2000.
- [27] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009.
- [28] R. P. Power, "The dominance of touch by vision: Sometimes incomplete," *Perception*, vol. 9, no. 4, pp. 457–466, Aug. 1980.
- [29] R. B. Welch and D. H. Warren, "Immediate perceptual response to intersensory discrepancy," *Psychol. Bull.*, vol. 88, no. 3, pp. 638–667, 1980.
- [30] J. F. Hayes, *Modeling and Analysis of Telecommunication Networks*. Hoboken, NJ, USA: Wiley, 2004.
- [31] J.-P. Bresciani, F. Dammeier, and M. O. Ernst, "Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events," *Brain Res. Bull.*, vol. 75, no. 6, pp. 753–760, Apr. 2008.
- [32] J.-P. Bresciani, M. Ernst, K. Drewing, G. Bouyer, V. Maury, and A. Kheddar, "Feeling what you hear: Auditory signals can modulate tactile tap perception," *Exp. Brain Res.*, vol. 162, pp. 80–172, May 2005.
- [33] K. Brunnström, E. Dima, T. Qureshi, M. Johanson, M. Andersson, and M. Sjöström, "Latency impact on quality of experience in a virtual reality simulator for remote control of machines," *Signal Process., Image Commun.*, vol. 89, Nov. 2020, Art. no. 116005.
- [34] C. Chaccour, R. Amer, B. Zhou, and W. Saad, "On the reliability of wireless virtual reality at terahertz (THz) frequencies," in *Proc. 10th IFIP Int. Conf. New Technol., Mobility Secur. (NTMS)*, Jun. 2019, pp. 1–5.
- [35] R. Fantacci and B. Picano, "End-to-end delay bound for wireless uVR services over 6G terahertz communications," *IEEE Internet Things J.*, early access, May 4, 2021, doi: [10.1109/JIOT.2021.3077497](https://doi.org/10.1109/JIOT.2021.3077497).
- [36] C. Chaccour, M. Soorki, W. Saad, M. Bennis, and P. Popovski, "Can terahertz provide high-rate reliable low latency communications for wireless VR?" 2020, *arXiv:2005.00536*.



BENEDETTA PICANO (Member, IEEE) received the B.S. degree in computer science and the M.Sc. degree in computer engineering and the Ph.D. degree in information engineering from the University of Florence. She was a Visiting Researcher with the University of Houston. Her research interests include matching theory, nonlinear time series analysis, resource allocation in edge and fog computing infrastructures, and machine learning.

• • •