Multi-Modal Signal Processing and Application in Communication

ISMAIL OLANIYI MURAINA[1], MOSES ADEOLU AGOI[2],
OLAYEMI MUYIDEEN ADESANYA[3], ABDULHAMEED OLUWATOSIN AMAO[4]

[1,2,3&4]Department of Computer Science, College of Information and Technology Education,
Lagos State University of Education, Lagos NIGERIA

Abstract: Multi-modal signal processing has emerged as a central and important technology in modern communication systems, blending information from various signal sources to improve communication efficiency, reliability, and accuracy. This paper explores the foundations of multi-modal signal processing, its significance in communication, and its practical applications. The study employs a descriptive research design using a purposive sampling technique to select respondents whose work is closely related to communications. 211 participants formed the population sample for the study with the survey instrument designed and administered via Google Forms which took 4 weeks. The validation of the instrument was done with the help of engineering colleagues while the reliability index was obtained as 0.789. The analysis was done using the SPSS 23.0 version. The study provides a comprehensive analysis of the methodologies involved, including data fusion, feature extraction, and machine learning algorithms tailored for multi-modal communication. Various application domains, such as audio-visual communication, biometric authentication, and human-computer interaction, are discussed to highlight the benefits of integrating different signal modalities. The study contributes to the understanding of multi-modal systems and their role in shaping the future of communication technologies.

Keywords: Multimodal Signals, Signal Processing, Communication, Single-Mode Signals, Multi-Dimensional Structure

Received: April 9, 2024. Revised: October 21, 2024. Accepted: November 7, 2024. Published: December 9, 2024.

1. Introduction

Over the past few decades, communication systems have rapidly expanded, requiring sophisticated signal processing algorithms to handle the growing volume and complexity of data. Single-mode signals, such as audio or visual signals alone, are frequently the subject of traditional signal processing. Real-world communication, on the other hand, is by its very nature multi-modal, combining signals from several modalities and sources. The integration and analysis of various signal types to produce a single, improved representation of the data being processed or sent is known as multi-modal signal processing. It is possible to define a framework for multimodal human-computer interfaces as one that combines media, styles, channels, and modes. While multimodality has been debated for some time in science and research, the computer science community is still learning how to create reliable and well-integrated multimodal systems. According to [1] , the proposed field of research is broad and encompasses a wide range of academic fields, including computer science, engineering, linguistics, cognitive sciences, human-computer interactions, and psychology. The field of multimodal signal processing is a crucial area of research and development that involves the processing of signals and the integration of data from multiple modalities, such as speech, vision, language, and text. This process greatly improves the comprehension, modelling, and functionality of systems or devices that facilitate human-computer interaction and human-to-human communication.

Extensive heterogeneous network installations generate massive volumes of data with high volume, high variety, high velocity, and high veracity characteristics. These data, often known as multimodal big data, present significant hurdles to conventional data fusion techniques and contain a wealth of cross- and intermodality information [2] . Living in the present era seems advantageous and handy with the advancement of mobile technologies. Numerous sensors and applications installed on mobile devices enable and continuously alter people's daily lives, including how they interact with one another, shop, and handle financial transactions [3] .

A brief segment of multimode fibre that has been fusion-spliced between two SMS fibres makes up a single-mode-multimode-single-mode (SMS) fibre structure [4] ,[5] , [6] . According to the study by [7] , biosensors have special opportunities to gather more data using a single sensing platform when EC approaches are integrated with varied modes/signals (such as light, magnetic, and thermal signals, etc.). The likelihood of false positives or negatives can be reduced and the detection accuracy can be further increased by connecting numerous signals or processing them logically. [8] produced multi-band/multi-mode radar signals for a variety of uses by utilizing the multidimensional properties of microwave photonics.

[9] defines multi-mode as a multi-dimensional structure that resembles a tensor. Accordingly, every mode in the multi-mode data has unique semantic and statistical characteristics. Intrinsic multi-mode properties of video data include cross-frame (temporal) mode and within-frame mode. In [33] . Provided an explanation of the mode-division multiplexing technology and designed a multi-mode coupler consisting of silicon-integrated multistage waveguide tapers and tapered few-mode fibre. This allowed for the direct coupling of high-order waveguide modes and high-order fibre modes with low crosstalk and high coupling efficiency. The good performance attained allowed for effective capacity scaling in fibre-chip-fibre optical interconnects and optical communication systems, opening up new avenues for a variety of mode-division multiplexing applications.

A multi-input multiple-output (MIMO) antenna array system that offers enhanced radiation diversity for multi-standard/multi-mode 5G communications was introduced in the study by [10] . The newly presented MIMO design, which has an overall size of 75 × 150 (mm2), consists of four pairs of miniature self-complementary antennas (SCAs) fed by pairs of independently connected structures that are symmetrically situated at the edge corners of the smartphone mainboard. Thus, for future multi-mode 5G cellular applications, antenna system design with enhanced radiation and multi-standard operation is a good contender.

The effective use of Yb doped multi-mode fibre amplifiers based pumped laser is simulated with cutting edge light sources in high speed modulated optical communication systems was the focus of [10] research. It is made clear how much total light power is measured using index multimode fibres with dual drive MZM measurements based on various baseband modulation voltages. Measure index multimode fibre mode field intensity distribution is simulated as a three-dimensional graph. The total electrical base ban power form is illustrated using a PIN photo-detector-based system with different modulation voltages and dual drive MZM measurements.

An examination of a solid-state transceiver-based 220 GHz multicarrier high-speed communication system was presented by [11] . The suggested method provides many signal carriers in the microwave band and converts them to a 220 GHz channel, thereby reducing the need for a high sampling rate analog-to-digital converter (ADC). A pair of 220 GHz solid-state transceivers with two signal carriers and two basebands for four GSPS ADCs make up the system. It has successfully transmitted real-time signals at a rate of 12.8 Gbps across a 20-meter distance using 16QAM modulation without the need for any additional test devices or supporting equipment. The baseband method ensures the viability of long-distance transmission applications by resolving the frequency difference issue caused by non-coherent structures.

[12] talked about the merging of data from several modalities, starting at the sensor level and moving on to the semantic level. Engineering signal processing research has long focused on sensor fusion, frequently for tracking things from satellites and aeroplanes. Physical occurrences are captured, interpreted, described, and altered by signal processing. Among the signal-processing languages they utilize to represent, evaluate, and extract meaningful information from real-world phenomena are statistics, probability, and stochastic processes [13] .

Fig. 1: Multimodal Communication

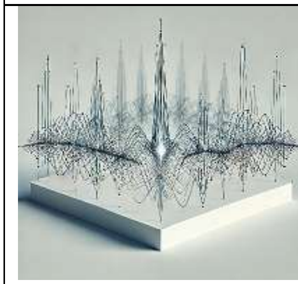 Single-mode signals	 Multi-mode Signals

Fig. 2: Comparison of Single Mode Signals and Multi-Mode Signals

The ability to merge auditory, visual, textual, and sensory signals has far-reaching ramifications for many industries, including telecommunications, healthcare, entertainment, and artificial intelligence. The goal of this study is to investigate the mechanics, algorithms, and practical applications of MSP, focusing on how multi-modal techniques might overcome the constraints of single-modal systems.

Fig. 3: The image combining audio, visual, textual, and sensory signals in a futuristic, abstract style

Objectives of the Paper

This paper aims to:

1. Provide an overview of the fundamental principles of multi-modal signal processing.

2. Analyze the techniques and methodologies used in multi-modal data fusion and feature extraction.

3. Explore the applications of multi-modal signal processing in communication systems.

4. Highlight the challenges and future research directions in the field of Multi-Signal Processing.

Research questions

1. How can multi-modal signal processing techniques improve the accuracy and efficiency of communication systems in noisy environments?
2. What are the key challenges and opportunities in integrating multi-modal signal processing for real-time communication across diverse platforms and networks?
3. To what extent does the fusion of audio, visual, and haptic signals in multi-modal processing enhance the quality of user experience in modern communication applications?

2. Literature Review

Motivation for Multi-Modal Signal Processing

Traditional communication systems use single-mode transmissions, which are frequently prone to errors caused by noise, distortions, or missing data. Multi-modal systems, on the other hand, make use of complementing information from several sources, increasing robustness and tolerance to noise while also boosting communication dependability. The integration of several modalities contributes to a more thorough understanding of the conveyed message, resulting in improved accuracy in tasks like speech recognition, facial recognition, and natural language understanding. According to [14] , in multimodal techniques, individual modalities are frequently integrated via simple fusion or directly fused with deep learning networks at the feature level. Numerous emotion recognition approaches have been proposed, most of which focus on visual, acoustic or psychophysiological information individually [14] .

Fundamentals of Multi-Modal Signal Processing

Multi-modal signal processing is the collecting, integration, and analysis of signals from several sources, including audio, video, and textual data. The fundamental goal is to integrate these signals in a meaningful fashion, hence improving the communication system's ability to read, transmit, and respond to information. Data fusion, feature extraction, and machine learning methods tailored to multi-modal data are among the most important techniques. A multimodal system allows for communication across various modalities or channels. In general, multimodal systems are thought to use concurrent processing and combine various, possibly asynchronous, input streams [1] .

<u>Data Fusion</u>

Fig. 4: An abstract artistic representation of data fusion, capturing the merging of multiple data sources into a unified form

The practice of merging data from several sources to provide more precise, trustworthy, and thorough information is known as data fusion. There are several data fusion levels, such as:
Feature-level fusion: A joint feature vector, which may be utilized for classification or decision-making, is created by combining extracted features from several modalities.
Decision-level fusion: The final decisions are integrated after each modality is processed separately.
Combining feature-level and decision-level fusion into one process is known as hybrid fusion.

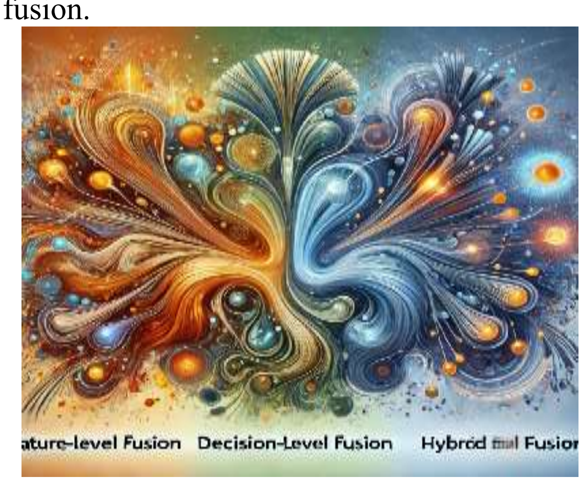

Fig. 5: An artistic illustration of Feature-level fusion, Decision-level fusion, and Hybrid fusion

Data fusion is a popular approach to handling incomplete raw data to capture trustworthy, valuable, and accurate information. When compared to various classical probabilistic data fusion techniques, machine learning—which automatically learns from prior experiences without explicitly programming—remarkably renovates fusion techniques by providing powerful computational and predictive ability [15] . Data fusion plays a crucial role in assisting AI in making sense of massive amounts of data at the network edge [16] .

A thorough overview of the history of data fusion and machine learning, including definitions, applications, architectures, processes, and common techniques, is given in [15] extensive survey on machine learning-based data fusion methods. Following this, they propose several requirements and use them as standards to assess and appraise the effectiveness of current machine learning-based fusion methods.

Data fusion and artificial intelligence (AI) at the edge were the main topics of [16] . They proposed a framework for data fusion and AI processing at the edge and then gave a comparative discussion of various data fusion and AI models and architectures. They also covered multiple levels of fusion and different types of AI, as well as how different types of AI align with different levels of fusion. Finally, they highlighted the advantages of combining data fusion and AI at the edge. The findings showed that combining data fusion and AI can boost speed by 9.8× and reduce energy consumption by up to 88.5% when compared to AI without data fusion.

To give readers, irrespective of their original community, an understanding of the principles of the multimodal deep learning fusion method and to inspire the development of new multimodal data fusion techniques of deep learning, Jing et al. (2020) offered a survey on deep learning for multimodal data fusion.

Feature Extraction and Dimensionality Reduction

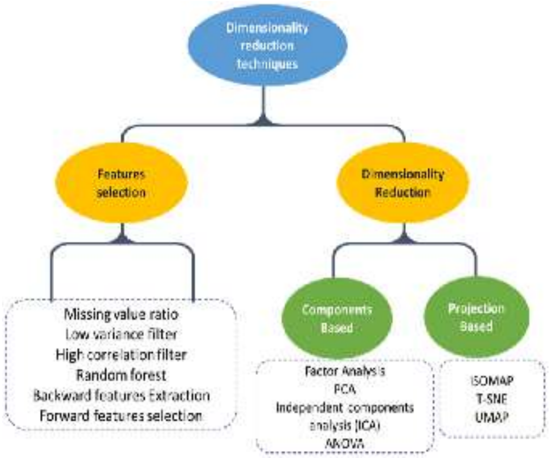

Fig. 6: Dimensionality Reduction Feature

Since multi-modal data is frequently high-dimensional, feature extraction and dimensionality reduction approaches are required. These methods keep the most important information intact while simplifying the data. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and different deep learning techniques are popular techniques. One essential tool for evaluating and comprehending high-dimensional data is the application of dimensionality reduction techniques. These methods collect some interesting data properties, including covariance, input-output linkages, dynamical structure, and correlation between data sets [17] . Mapping a set of high-dimensional data features onto low-dimensional data is the process of dimensionality reduction.

Dimensionality reduction (DR) is a pre-processing step that has been suggested and put into practice utilizing feature selection and extraction methods. Its goal is to increase learning feature accuracy and decrease training time by eliminating redundant features, and noisy, and irrelevant data. One Dimension Reduction technique that reduces calculation time for the learning process is Principal Component Analysis (PCA) [18] . [19] state that every data mining or machine learning (ML) task demands the use of increasingly effective techniques to get the required results, due to the dramatic growth in data dimensions. Dimensionality reduction is used as a pre-processing step to improve learning feature accuracy and save training time. It can remove noise and unnecessary data.

The two primary techniques used for dimensionality reduction (DR) are feature extraction (FE) and feature selection (FS). Because data is generated continually and at an ever-increasing rate, FS is regarded as a significant method. It may effectively minimize redundancy, remove irrelevant data, and improve result comprehensibility, among other serious dimensionality issues. Furthermore, FE deals with the issue of identifying the most unique, informative, and reduced set of features to improve the effectiveness of data processing and storage.

Machine Learning in Multi-Modal Signal Processing

Machine learning techniques, like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are often used in combination with multi-modal data to enhance performance in tasks like image recognition, speech recognition, and natural language processing. Machine learning algorithms are essential to processing multi-modal data because they allow systems to automatically learn patterns and relationships across different modalities.

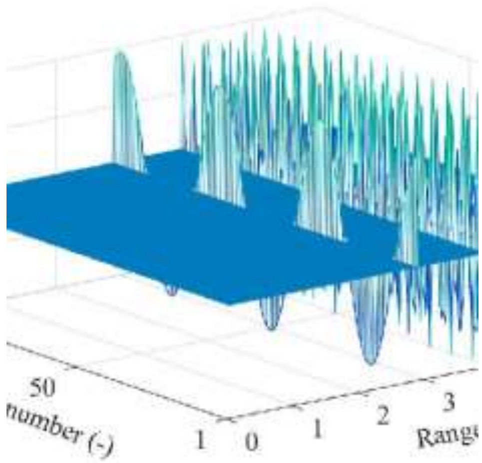

Fig. 7: Machine Learning in Multi-Modal Signal

Combining machine learning (ML) with heterogeneous data from several modalities to address important problems is the exciting field of multimodal machine learning (MML). Research projects often use data from one modality, like text, audio, photos, and signals. But nowadays, problems in the real world are critical, and solving them with numerous data modalities rather than just one might have a big impact on the outcome. By adjusting parameters, ML algorithms are crucial to the creation of MML models [20] .

The goal of the dynamic, multidisciplinary research field of multimodal machine learning is to create computer agents that are intelligent enough to comprehend, reason, and learn by combining various communicative modalities, such as linguistic, acoustic, visual, tactile, and physiological messages. Recent application domains have shown interest in text-to-image generation, embodied autonomous agents, multisensor fusion, and video understanding. In 2024, Paul, Amir, and Louis-Philippe said that while unimodal learning has advanced significantly, it still does not fully capture the complexity of human learning. Multimodal learning aids in better understanding and analysis of situations in which multiple senses are involved in the processing of information [21] . Deep learning has been applied to a wide range of applications and has grown in popularity in recent years. The goal of multimodal deep learning (MMDL) is to create models that can process and link information using various modalities.

Applications of Multi-Modal Signal Processing in Communication

Multi-modal signal processing has numerous applications in communication, ranging from enhancing human-to-human interaction to enabling more intuitive human-computer interaction. Some key applications include:

Audio-Visual Communication

Fig. 8: Audio-visual Communication

Combining audio and visual signals in communications improves information delivery. For instance, video conferencing makes use of both visual and aural cues to produce a more interesting and productive communication environment. To improve comprehension and reaction accuracy, audio-visual integration is also utilized in systems such as interactive robots and virtual assistants. It has long been a practice to use audio-visual (AV) technologies for instruction and learning [22] .

Since deep learning began to be applied successfully, there has been a lot of interest in audio-visual learning, which aims to take advantage of the link between auditory and visual modalities. Researchers tend to exploit these two modalities to improve the performance of previously regarded single-modality activities or handle new tough challenges [23] . Complex human behaviour can be understood by studying physical features from multiple modalities; mainly facial, vocal and physical gestures [24]

Biometric Authentication

Fig. 9: Biometric Sample types

To increase the accuracy of identification, biometric systems integrate multiple biometric signals—such as voice recognition, fingerprint scanning, and facial recognition—using multi-modal signal processing. These systems are commonly utilized in security applications, including protecting sensitive data and unlocking devices. Biometrics is the automated process of identifying an individual based on their biological and behavioural characteristics [25] Technologies for biometric identity authentication are widely used nowadays.

According to [26] , these systems are used in businesses, controlled-access facilities, regular people's smartphones, and online applications. [3] state that biometrics, which are characteristics shared by all humans, are used to uniquely identify each individual. Creating protections against unauthorized access and authentication is essential to system security. The detection of impersonation attacks is a problem for current user authentication mechanisms, making systems open to abuse. Continuous multimodal biometric authentication (CMBA) systems have been recommended as a dependable option in many research projects [27] .

Human-Computer Interaction (HCI)

Fig. 10: Sample of HCI

Enhancing human-computer interaction requires multimodal signal processing. HCI systems can interpret user commands more accurately by integrating speech, gestures, and visual clues, resulting in a more fluid and natural interaction. Multi-modal signals play a major role in virtual reality and augmented reality technologies' ability to produce immersive experiences. The Internet of Things (IoT), mobile and cloud computing, and other emerging technologies are requiring more attention from human-computer interaction (HCI) experts in terms of systems interface design. [28] noted that these days, modern information systems (emerging technologies) are increasingly becoming an integral part of our daily lives and have begun to pose a serious challenge for HCI professionals.

Children, the elderly, and those with disabilities or disorders make up a large portion of the mobile platform user base these days, and they all have different demands that must be met by an efficient user interface that can work with them while they are on the go, at any time, or anyplace [28] . The significance of taking values into account while designing technology related to human-computer interaction (HCI) is becoming more widely recognized (Eriksson et al., 2022; Choudhury et al., 2020).

<u>Medical Communication Systems</u>

Fig. 11: Sample of Medical Communication Systems

Multi-modal signal processing can be used in the medical field to combine signals from sensor data, patient records, and medical imaging to enhance treatment results and diagnostic precision. For example, merging information from CT scans, MRIs, and patient histories enables doctors to make better decisions. These days, developing smart healthcare systems is a trendy trend, and the foundation of these systems is wireless communication. All healthcare facilities need to be outfitted with cutting-edge technology to offer patients diagnoses, treatments, and a variety of other health services both on-site and through remote providers. During disasters, a rapid response unit is necessary to manage the influx of patients and inquiries. [31]
[32] found that to treat potentially fatal disorders, quick access to emergency medical communication centres (EMCCs) is essential. Even with the markedly increased activity and the challenges associated with mobilizing more staff resources, the public must continue to have immediate access to EMCCs in the event of a disease epidemic. In healthcare systems, multi-modal signal processing facilitates communication between patients, doctors, and other stakeholders [33] . Individuals who are mentally, speech, or hearing impaired need specific communication support, particularly when it comes to health issues [34] .

Methodology

The study employed a descriptive research design focusing on Communication Engineers, Researchers in Signal Processing and Communication Technologies, Telecommunication Professionals, Graduate and Postgraduate Students in Electrical Engineering or Communication Systems, Industry Experts in AI, Machine Learning, and Signal Processing and Government and Regulatory Bodies in Telecommunications as target population. 211 formed the population sample for the study selected purposively with the survey instrument designed and administered via Google Forms which took 4 weeks. The validation of the instrument was done with the help of engineering colleagues while the reliability index was obtained as 0.789. The analysis was done using the SPSS 23.0 version.

3. Analysis and Results

Demographic Information

Table 1: Participants involved in the study

Participants	**Roles**	**Frequency**
Communication Engineers	Experts in signal processing, data transmission, and telecommunications who can provide insights into real-world applications	**37**
Researchers in Signal Processing and Communication Technologies	Academics or professionals working in fields related to signal processing, such	**36**

	as multimedia communication, audio-visual processing, or sensor fusion	
Telecommunication Professionals	Engineers and professionals working with telecom companies who implement multi-modal signal processing in network infrastructure	**36**
Graduate and Postgraduate Students in Electrical Engineering or Communication Systems	Students studying or specializing in signal processing or communication technologies, as can participate in experimental studies or simulations	**36**
Industry Experts in AI, Machine Learning, and Signal Processing	Professionals using AI and machine learning techniques to improve multi-modal signal processing applications, such as speech recognition, natural language processing, or image recognition	**36**
Government and Regulatory Bodies in Telecommunications	Agencies responsible for setting standards or overseeing communication infrastructure	**36**

The study's participants are listed in Table 1. It shows that there were 37 communication engineers (17.54%), while the following groups also had similar numbers of 36 each: Telecommunication Professionals, Industry Experts in AI, Machine Learning, and Signal Processing, Government and Regulatory Bodies in Telecommunications, Graduate and Postgraduate Students in Electrical Engineering or Communication Systems, and Researchers in Signal Processing and Communication Technologies.

RQ1: How can multi-modal signal processing techniques improve the accuracy and efficiency of communication systems in noisy environments?

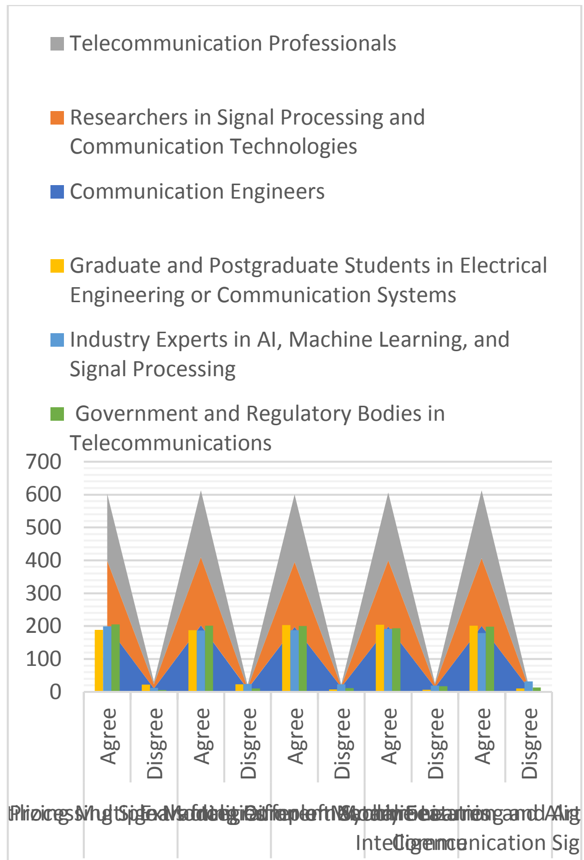

Fig.12: Participants' response on how multi-modal signal processing techniques can improve the accuracy and efficiency of communication systems in noisy environments

The majority of respondents in Fig. 1 concurred that improving the accuracy and effectiveness of communication systems in noisy environments can be achieved by using multiple modalities, processing signals from various modalities, extracting complementary features, integrating machine learning and artificial intelligence, and synchronizing and aligning communication signals.

RQ2: What are the key challenges and opportunities in integrating multi-modal signal processing for real-time communication across diverse platforms and networks?

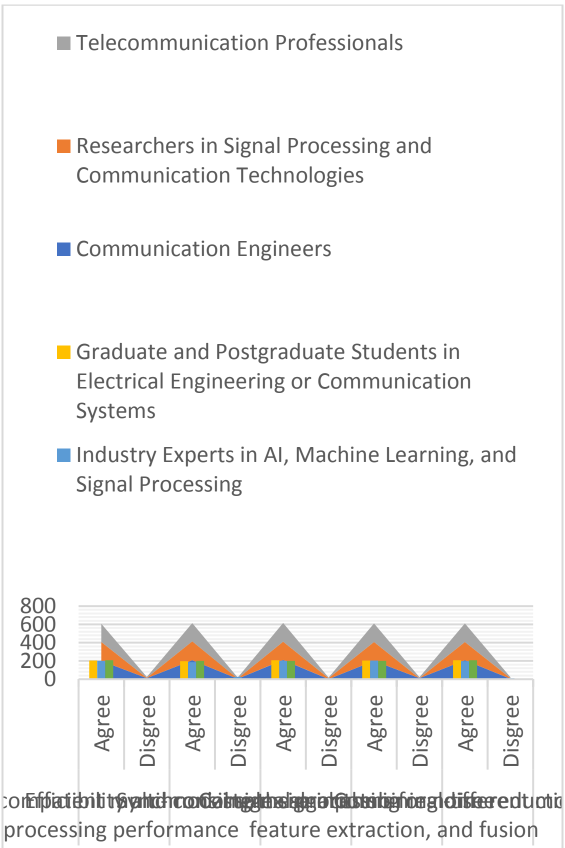

Fig.13: Participants' response on the key challenges and opportunities in integrating multi-modal signal processing for real-time communication across diverse platforms and networks

Similarly, Fig. 2 demonstrates that most respondents concurred that the following factors are important for integrating multi-modal signal processing for real-time communication across various platforms and networks: efficient multi-modal signal processing, synchronizing inputs in real-time, complex algorithms for noise reduction, feature extraction, and fusion, and combining different modalities.

RQ3: To what extent does the fusion of audio, visual, and haptic signals in multi-modal processing enhance the quality of user experience in modern communication applications?

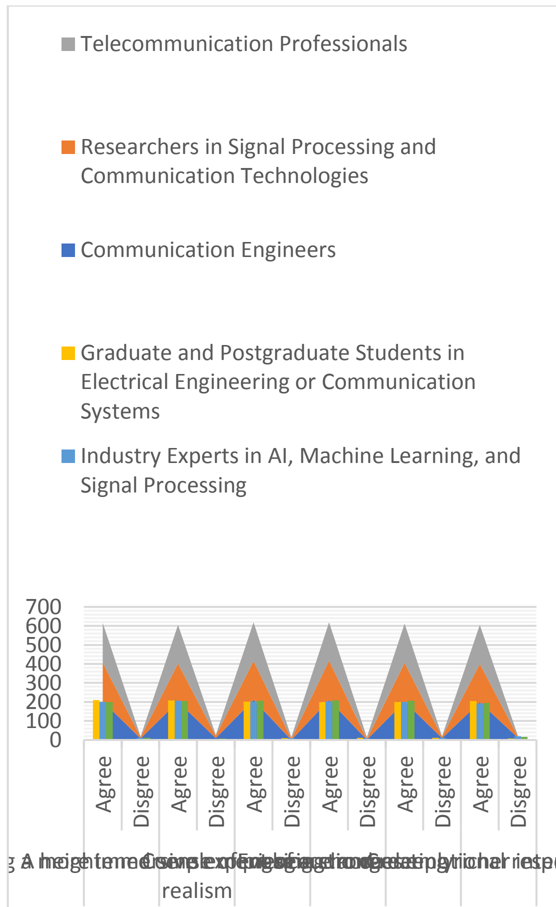

Fig.14: Participants' response on the extent to which the fusion of audio, visual, and haptic signals in multi-modal processing can enhance the quality of user experience in modern communication applications

The majority of respondents also concurred, as shown in Fig. 3, that the fusion of audio, visual, and haptic signals in multi-modal processing can improve the quality of user experience in contemporary communication applications by producing a more immersive experience, a heightened sense of presence and realism, a complement of audio cues, engaging more deeply, evoking stronger emotional responses, and creating richer interactions.

4. Discussion

The findings show that communication systems can reduce dependency on a single signal source by utilising multiple modalities, such as speech and visual signals. In noisy environments, multi-modal systems can extract complementary features that improve overall communication accuracy. The integration of machine learning and artificial intelligence (AI) can help filter out noise by recognizing patterns across modalities. Signal enhancement techniques, such as beamforming in audio-visual systems, focus on the desired signal while suppressing background noise. Furthermore, multi-modal signal processing systems must handle distinct bandwidths, latencies, and error rates across different networks. Efficient multi-modal signal processing must handle these variations without compromising the quality of communication. Real-time communication demands fast and efficient signal processing algorithms, which must be optimized for each device's processing power.

Scaling efficiently is crucial for multi-modal signal processing systems, as they must remain robust as communication networks grow while maintaining quality of service. Ensuring secure communication across different platforms and networks requires adhering to various security standards. The integration of multi-modal signals enhances the richness of communication, enabling more immersive experiences, such as virtual or augmented reality. AI-driven multi-modal systems can dynamically adapt to user preferences, environments, and devices. The rollout of 5G and emerging network technologies provides higher bandwidth and lower latency, enabling smoother and more reliable real-time multi-modal communication. Combining multi-modal signal processing with edge computing can improve latency and efficiency by processing data closer to the source. AI and machine learning can be used to improve signal processing tasks such as noise suppression, predictive analytics for network optimization, and automatic translation across multiple languages in real-time. The ongoing development of global standards for interoperability across platforms is creating opportunities for seamless integration of multi-modal communication technologies. Advancements in APIs and middleware can make multi-modal systems integrate diverse communication platforms and networks more easily, enabling seamless real-time processing across different ecosystems.

The integration of multiple sensory modalities, including audio, visual, and haptic, creates a more immersive experience, mimicking how humans naturally interact with the world. In virtual or augmented reality applications, visual elements are combined with spatial audio and haptic feedback to make the virtual environment feel more tangible and realistic. This combination enhances comprehension and engagement, making users feel more involved in the process. Haptic feedback in combination with audio and visual cues can evoke stronger emotional responses, making users feel physically involved in the interaction. When combined, users can process information faster and with less effort, making interactions more intuitive. Multi-modal systems can also enhance accessibility for users with disabilities, ensuring that all users can participate fully regardless of sensory limitations. In VR and AR environments, fusing visual, audio, and haptic signals creates richer interactions, making remote interactions more engaging and responsive. Multi-modal communication systems are vital in assistive technologies, helping individuals with disabilities interact with devices more effectively.

5. Challenges and Future Research Directions

Despite its numerous benefits, multi-modal signal processing presents several challenges that need to be addressed. These include:

Synchronization issues: Combining signals from different sources can be challenging due to the need for accurate synchronization.

Data heterogeneity: Different modalities often have varying data structures and formats, complicating the fusion process.

Computational complexity: Processing high-dimensional multi-modal data requires significant computational resources.

Future research in MSP is likely to focus on developing more efficient algorithms for real-time data fusion, improving the scalability of multi-modal systems, and exploring new applications in emerging fields like 6G communication and autonomous vehicles.

6. Conclusion

Through the integration of data from several sources, multi-modal signal processing improves accuracy, dependability, and efficiency in communication. Its applications cover a wide range of sectors, including healthcare and telecommunications, and its influence on communication technology is only increasing. The evolution of machine learning, data fusion, and signal processing techniques will significantly impact the future of communication, with multi-modal systems being a key player. Techniques for multi-modal signal processing improve the efficiency of communication systems in noisy settings.

These methods enable more precise and effective communication even in difficult situations by utilizing complementing data from several modalities, adding AI for noise reduction, and enhancing error correction and synchronization. Although there are some obstacles to overcome, including latency, network unpredictability, data fusion, and security issues, integrating multi-modal signal processing into real-time communication across various platforms and networks also presents several substantial opportunities.

Enhancing user experiences, facilitating more contextual and immersive communication, utilizing edge computing and AI, and promoting innovation in sectors like healthcare, education, and smart cities are a few of these. The possibilities will keep expanding as technology develops, especially with the rollout of 5G and the acceptance of new standards, enabling more intelligent, reliable, and seamless real-time multi-modal communication.

A more immersive, interesting, and approachable interaction is made possible by the combination of audio, visual, and haptic signals in multi-modal processing. This multisensory approach not only improves perception and realism but also lowers cognitive load, boosts emotional engagement, and encourages inclusivity. As technology advances, the integration of these modalities will probably become even more complex, spurring more innovation in contemporary communication applications.

References

[1] Ledin, Per & Machin, David. (2020). Introduction to Multimodal Analysis: Second Edition. 10.5040/9781350069176

[2] Jing Gao, Peng Li, Zhikui Chen, Jianing Zhang; A Survey on Deep

Learning for Multimodal Data Fusion. Neural Comput 2020; 32 (5): 829–864. doi: https://doi.org/10.1162/neco_a_01273

[3] Silasai, O., & Khowfa, W. (2020). The study on using biometric authentication on mobile devices. NU Int. J. Sci, 17, 90-110.

[4] Wu, Qiang & Qu, Yuwei & Liu, Juan & Yuan, Jin-Hui & Wan, Sheng-Peng & Wu, Tao & He, Xing-Dao & Liu, Bin & Liu, Dejun & Ma, Youqiao & Semenova, Yuliya & Wang, Pengfei & Xin, Xiangjun & Farrell, Gerald. (2020). Singlemode-Multimode-Singlemode Fiber Structures for Sensing Applications – A Review. IEEE Sensors Journal. PP. 1-1. 10.1109/JSEN.2020.3039912.

[5] Girsang, Liliana; Napitupulu, Angelin; Simorangkir, Irmalia & Sitompul, Dahlan. (2021). In-depth Study of Single Mode Optical Fibre. https://www.researchgate.net/publication/357364138_Indepth_Study_of_Single_mode_Optical_Fibre

[6] Razzaq Salih, Aqeel. (2020). Design of Single Mode Fiber for Optical Communications. Ibn AL- Haitham Journal For Pure and Applied Science. 33. 40-47. 10.30526/33.1.2373.

[7] Qingzhi Han, Haimei Wang, John Wang (2024). Multi-Mode/Signal Biosensors: Electrochemical Integrated Sensing Techniques.https://doi.org/10.1002/adfm.202403122

[8] Ragheb, A., Esmail, M.A., Seleem, H. et al. Photonics-based multi-band/multi-mode radar signal generation. Photon Netw Commun 39, 91–101 (2020). https://doi.org/10.1007/s11107-019-00859-7

[9] Haohua Zhao, Weichen Xue, Xiaobo Li, Zhangxuan Gu, Li Niu, Liqing Zhang (2020). Multi-mode neural network for human action recognition. https://doi.org/10.1049/iet-cvi.2019.0761

[10] Tamilselvi, M., Kumar, A., Ganapathy, D., Devi, P., Ramachandran, T., Zahran, A. & Yesudasu, V. (2023). Efficient employment of Yb doped multi-mode fibre amplifiers based pumped laser with advanced light sources in high speed modulated optical communication systems. Journal of Optical Communications. https://doi.org/10.1515/joc-2023-0206

[11] Niu, Z., Zhang, B., Wang, J., Liu, K., Chen, Z., Yang, K., ... & Liu, Y. (2020). The research on 220GHz multicarrier high-speed communication system. China Communications, 17(3), 131-139.

[12] Oviatt, S., Cohen, P.R. (2015). Multimodal Signal Processing, Fusion, and Architectures. In: The Paradigm Shift to Multimodality in Contemporary Computer Interfaces. Synthesis Lectures on Human-Centered Informatics. Springer, Cham. https://doi.org/10.1007/978-3-031-02213-5_7

[13] Sun, X., Tian, Y., Lu, W. et al. From single- to multi-modal remote sensing imagery interpretation: a survey and taxonomy. Sci. China Inf. Sci. 66, 140301 (2023). https://doi.org/10.1007/s11432-022-3588-0

[14] Wang, Z., Zhou, X., Wang, W. et al. Emotion recognition using multimodal deep learning in multiple

psychophysiological signals and video. Int. J. Mach. Learn. & Cyber. 11, 923–934 (2020). https://doi.org/10.1007/s13042-019-01056-8

[15] Meng, Tong & Jing, Xuyang & Pedrycz, Witold. (2019). A Survey on Machine Learning for Data Fusion. Information Fusion. 57. 10.1016/j.inffus.2019.12.001.

[16] Munir, Arslan; Blasch, Erik; Kwon, Jisu; Kong, Joonho & Aved, Alexander. (2021). Artificial Intelligence and Data Fusion at the Edge. IEEE Aerospace and Electronic Systems Magazine. 36. 62-78. 10.1109/MAES.2020.3043072.

[17] Ahmad, Noor & Nassif, Ali. (2022). Dimensionality Reduction: Challenges and Solutions. ITM Web of Conferences. 43. 01017. 10.1051/itmconf/20224301017.

[18] Sarveshwaran, Velliangiri; Sevugan, Alagumuthukrishnan & Swamidason, Iwin Thanakumar Joseph. (2019). A Review of Dimensionality Reduction Techniques for Efficient Computation. Procedia Computer Science. 165. 104-111. 10.1016/j.procs.2020.01.079.

[19] Zebari, Rizgar & Abdulazeez, Adnan & Zeebaree, Diyar & Zebari, Dilovan & Saeed, Jwan. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. Journal of Applied Science and Technology Trends. 1. 56-70. 10.38094/jastt1224.

[20] Barua, Arnab; Ahmed, Mobyen & Begum, Shahina. (2023). A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3243854.

[21] Summaira Jabeen, Xi Li, Muhammad Shoib Amin, Omar Bourahla, Songyuan Li, and Abdul Jabbar. 2023. A Review on Methods and Applications in Multimodal Deep Learning. ACM Trans. Multimedia Comput. Commun. Appl. 19, 2s, Article 76, 41 pages. https://doi.org/10.1145/3545572

[22] Owusu, A. (2020). The impact of audio-visual technologies on university teaching and learning in a developing economy. South African Journal of Information Management, 22(1), 1-9.

[23] Zhu, H., Luo, M. D., Wang, R., Zheng, A. H., & He, R. (2021). Deep audio-visual learning: A survey. International Journal of Automation and Computing, 18(3), 351-376

[24] Schoneveld, L., Othmani, A., & Abdelkawy, H. (2021). Leveraging recent advances in deep learning for audio-visual emotion recognition. Pattern Recognition Letters, 146, 1-7.

[25] Sarkar, A., & Singh, B. K. (2020). A review of performance, security and various biometric template protection schemes for biometric authentication systems. Multimedia Tools and Applications, 79(37), 27721-27776.

[26] Kuznetsov, A., Oleshko, I., Chernov, K., Bagmut, M., & Smirnova, T. (2019, April). Biometric authentication using convolutional neural networks. In Conference on Mathematical Control Theory (pp. 85-98). Cham: Springer International Publishing.

[27] Ryu, R., Yeom, S., Kim, S. H., & Herbert, D. (2021). Continuous multimodal biometric authentication

schemes: a systematic review. IEEE Access, 9, 34541-34557.

[28] Mohammed, Y. B., & Karagozlu, D. (2021). A review of human-computer interaction design approaches towards information systems development. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 12(1), 229-250. DOI: https://doi.org/10.18662/brain/12.1/180

[29] Eriksson, E., Nilsson, E. M., Hansen, A. M., & Bekker, T. (2022). Teaching for values in human-computer interaction. Frontiers in Computer Science, 4, 830736. https://doi.org/10.3389/fcomp.2022.830736

[30] Choudhury, M. D., Lee, M. K., Zhu, H., & Shamma, D. A. (2020). Introduction to this special issue on unifying human-computer interaction and artificial intelligence. Human–Computer Interaction, 35(5–6), 355–361. https://doi.org/10.1080/07370024.2020.1744146

[31] Janjua, M. B., Duranay, A. E., & Arslan, H. (2020). Role of wireless communication in the healthcare system to cater to disaster situations under 6G vision. Frontiers in Communications and Networks, 1, 610879. https://doi.org/10.3389/frcmn.2020.610879

[32] Penverne, Y., Leclere, B., Labady, J., Berthier, F., Jenvrin, J., Javaudin, F., & Montassier, E. (2020). Impact of two-level filtering on emergency medical communication centre triage during the COVID-19 pandemic: an uncontrolled before-after study. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 28, 1-6. https://doi.org/10.1186/s13049-020-00775-0

[33] Xiaoping Cao, Kang Li, Yuanjian Wan & Jian Wang (2022). "Efficient mode coupling between a few-mode fibre and multi-mode photonic chip with low crosstalk," Opt. Express 30, 22637-22648 (2022), https://opg.optica.org/oe/fulltext.cfm?uri=oe-30-13-22637&id=476758

[34] Wołk, K., Wołk, A., & Glinkowski, W. (2017). A Cross-Lingual Mobile Medical Communication System Prototype for Foreigners and Subjects with Speech, Hearing, and Mental Disabilities Based on Pictograms. Computational and mathematical methods in medicine, 2017(1), 4306416. https://doi.org/10.1155/2017/4306416