International Journal of Business Analytics
Volume 12 • Issue 1 • January-December 2025

Music Emotion Recognition-Based Business-Oriented Visualization Framework Using AI-driven Serverless Cloud Computing

Muhammed Golec
https://orcid.org/0000-0003-0146-9735
Queen Mary University of London, UK

Lifeng Zhu
Queen Mary University of London, UK

Emir Sahin Hatay
https://orcid.org/0009-0003-0240-9854
University of Essex, UK

Han Wang
https://orcid.org/0009-0002-4984-9307
Queen Mary University of London, UK

Sukhpal Singh Gill
https://orcid.org/0000-0002-3913-0369
Queen Mary University of London, UK

ABSTRACT

This paper proposes a novel framework for a real-time music visualization system designed for the hearing-impaired, utilizing AI and serverless computing. The system converts audio signals into visual representations that capture both the physical and emotional aspects of music. A neural network-based Music Emotion Recognition (MER) model extracts emotional cues, which are integrated into the visualizations. The serverless computing ensures accessibility, while an account management system and comment collection system enable customization and regular retraining of the model for better accuracy. Results demonstrate the framework's effectiveness, highlighting the scalability and cost-efficiency of serverless computing. This work significantly advances music accessibility for the hearing-impaired, enhancing sensory experiences and promoting mental well- being. The MER model shows superior performance, with a 46.8% lower Root Mean Squared Error (RMSE) compared to other works targeting the same 10-second audio fragment length and a 13.5% higher Pearson's correlation coefficient (PCC) for 30-second fragments.

KEYWORDS

Real-Time Music Visualization, Hearing Impaired, Music Emotion Recognition, Neural Networks, Serverless Computing, Personalized Music Experience

DOI: 10.4018/IJBAN.373258

1. INTRODUCTION

The World Health Organization (WHO) estimates that around 1.3 billion people, or 16% of the global population, live with significant disabilities (World Health Organization [WHO], 2024). About 43 million of them are blind (International Agency for the Prevention of Blindness, 2020) and 430 million of them are deaf (World Health Organization [WHO], 2024). Facilities for the disabled are nowadays lively developed and extremely commonly deployed, bringing more convenience into their daily life, like blind sidewalk, braille button and sound traffic light. But physical difficulties aside, the mental health of the disabled has always been a significant problem that should not be ignored (Salako, I., 2017). Therefore, convenient facilities and equipment targeting their mortal life are also wealth to be attached importance to, like helping the blind to visit a museum or the deaf to attend a concert. Deafness and hearing loss are found worldwide in all regions and countries. Approximately one-fifth of the global population are living with hearing loss and 430 million of them are suffering disabling hearing loss (World Health Organization [WHO], 2024). It is expected that there could be over 700 million people with disabling hearing loss by 2050 (International Agency for the Prevention of Blindness, 2020). 34 million children have deafness or hearing loss and about 30% of people over 60 have hearing loss (World Health Organization [WHO], 2024). Besides, by the growth of age, the ability of hearing low magnitude or high frequency voice will degenerate significantly, (Oh, Lee, Park, Kim, Chung, Kim, & Yeo, 2014) which will also impact the ability to sense the music.

Employing medical technique is one of the solutions. Equipment like bone-anchored hearing devices (Baker, S., Centric, A., & Chennupati, S. K., 2015) can be deployed by some specific surgery procedures. But on one hand, implants have been proved not to be able to provide enough assist to the hearing impacted caused by all the reasons (Lupo, J. E., Biever, A., & Kelsall, D. C., 2020), on the other hand, the cost of implant surgery is not easy for every family to afford. Another family of solutions to help people with hearing loss to enjoy music is to transform the music into other communication modalities more familiar to deaf people, which shows more potential for higher availability and expansibility with the support of rapidly developing advancements.

Music information can be roughly separated into two properties, physical and perceptual. The physical part consists of amplitude, frequency spectrum and timbre, while the perceptual part represents the emotion it carries. To achieve the target of passing this information to the deaf, to extract these characteristics from the audio steam is the key. The most well-known physical feature representing system is pitch + magnitude + duration, which is utilized in the famous Musical Instrument Digital Interface (MIDI) filespec (MIDI Manufacturers Association., 2023). In the MER field, Russell (Russell, J. A., 1980) proposed circumplex model consists of two dimensions: valence and arousal, which is then adapted and improved by Thayer (Thayer, R., 1989), forming the widely used model of valence and arousal dimensions. Valence shows the pleasure level of the mood (displeasure to pleasure) and arousal shows the excited degree of it (clam to excited).

In the context of transforming audio into forms the deaf more familiar with, it has been proved that deaf people show better reaction on visual flashes than ones with regular hearing (Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K., 2015). Besides, it has also been proved that the deaf process visual and vibrotactile information utilizing the brain area that is normally used for hearing (Good, A., Reed, M. J., & Russo, F. A., 2014). Therefore, dynamic image and vibration are both good choices. To achieve the high availability to basically all the target population, responsive design in cloud computing technique will be employed to ensure that can be used on any kind of smart device in any situation as long as there is Internet (Bernacki, J., Błaz˙ejczyk, I., Indyka-Piasecka, A., Kopel, M., Kukla, E., & Trawiński, B., 2016). Considering the end carrier of this server is very likely to be a tablet or a cellphone, using dynamic image is the best option to implement. (Nanayakkara, S. C., 2008) and (Ho-Ching, F., McGrenere, J. S., & Graf, P. M., 2003) have suggested a music visualization alone is enough to provide musical experience to the users. To cover the information loss during the transformation process to the highest extent possible, merging MER into the audio processing is