Broadcasting and Hosting with Algorithm Assistance

Yuanchun Wu*

School of Broadcasting and Anchoring, Communication University of China, Beijing 100024, China

*2283969514@qq.com

Abstract. In the context of artificial intelligence and algorithmic influence on media ecology, what kind of impact will hosting communication face? When considering news production, news ecology, and news media in the era of intelligent media, it is necessary to weigh the pros and cons of artificial intelligence in language communication, allowing human value to return to language, and fully leveraging the vital role of spoken language in exchanging ideas and conveying emotions. Thanks to the assistance of algorithms, whether it's simulating human voice, recognizing voice quality, or classifying acoustic scenes, algorithms demonstrate their advantages in language communication, efficiently and effectively facilitating the dissemination of information.

Keywords: artificial intelligence, algorithmic, media, hosting communication, ideas, emotions, voice, information

1 Introduction

With the widespread application of artificial intelligence and Internet of Things (IoT) technology in the media, technological innovations represented by big data applications, robot journalism, drone gathering, and virtual reality are profoundly impacting and changing the existing media ecology. Technology is leading the media into the "era of intelligent media." Artificial intelligence technology has brought about a new mode of news production different from traditional media, and the transformation based on algorithmic perspectives has further promoted the personalized production of news. In the broader context of artificial intelligence and algorithms deeply influencing the news media ecology, it is necessary to consider the new changes in hosting communication and analyze the pros and cons of artificial intelligence in language communication, thereby exploring the value of hosting communication in the era of intelligent media. In occasions where strict requirements are placed on voice quality for broadcasting and hosting, there is a broad market demand for the analysis and selection of voice quality. Voice simulation, to a certain extent, reduces the workload of practitioners, allowing them to focus more on quality control. Therefore, the classification of acoustic scenes is also indispensable. In conclusion, it is necessary to conduct research and analysis on voice simulation, voice quality recognition, and acoustic scene classification.

E. P. H. Lau et al. (eds.), *Proceedings of the 2024 3rd International Conference on Information Economy, Data Modelling and Cloud Computing (ICIDC 2024)*, Advances in Computer Science Research 114,
https://doi.org/10.2991/978-94-6463-504-1_19

2 Artificial Intelligence Voice Simulation Enriches Hosting Communication in the Era of Intelligent Media[1][2]

The development of artificial intelligence technology has not only changed the forms and mechanisms of media, but also enriched the forms of hosting communication, particularly evident in the widespread application of AI voice simulation. The integrated development of "AI + media" has expanded the application scope of artificial intelligence in hosting communication, gradually transitioning from research applications in the audio broadcasting field to "audio + 3D animation." With the development of virtual image generation (P2A) and virtual image-driven technology, the voice, image, and functionality of virtual hosts continue to improve. At the same time, virtual hosts satisfy the need for interaction with the audience, allowing viewers to customize the appearance of virtual hosts according to their personal preferences, including clothing, styling, lighting, and background, thereby meeting the diverse needs of different audiences to a large extent. In the intelligent era where significant changes and influences have occurred in both "listening" and "speaking," diverse choices cater to the needs of different audiences, especially in the field of hosting communication. The unique expression and rich virtual images, supported by intelligent technology, are widely applied, providing a more diverse range of choices for the production and creation of communication information while achieving the integration of technology and art, demonstrating the "human-machine interaction" in hosting communication in the era of intelligent media.

The MOS value reflects the subjective clarity of speech quality. According to ITU-T G.107, the mapping relationship between parameter R and MOS values is defined as follows:

$$MOS = \begin{cases} 1, R < 0 \\ 1 + 0.035R + 7R(R-60)(100-R) \times 10^{-6}, 0 < R < 100 \\ 4.5, R > 100 \end{cases} \quad (1)$$

We can evaluate the quality of speech based on the above formula.

3 Algorithm-Based Voice Quality Recognition[3][4]

The information regarding a speaker's voice quality is embedded within the recognized information of the speaker. The analysis of voice quality falls under the branch of speaker recognition, thus research on speaker recognition is equally applicable to the analysis of voice quality in both theory and practice. Research on speaker recognition began as early as the 1930s, with early work focusing on human ear discernment experiments and exploring the possibilities of auditory recognition. As research methods and tools have improved, studies have progressed beyond simple human ear discernment.

The recognition model has evolved from single-template models to multi-template models, from template models to vector quantization models, Gaussian mixture models,

hidden Markov models, and then to artificial neural networks. The recognition environment has progressed from recognizing a small number of speakers in noise-free environments to recognizing a large number of speakers in complex environments. The recognition technology employed has developed from solely involving dynamic programming to encompassing statistical signal processing, vector quantization and coding, system theory and methods, optimal estimation theory, artificial neural networks, and grey system analysis across multiple disciplines.

Figure 1 illustrates the framework of a speaker voice quality recognition system. Similar to a speech recognition system, establishing and applying such a system can be divided into two stages: training and recognition. During the training stage, speakers utter several training sentences for each voice quality trait, upon which the system constructs templates or model parameters for each voice quality trait. In the recognition stage, the parameters extracted from the speech with the voice quality traits to be recognized are compared with the reference parameters or templates obtained during training, and a judgment is made based on certain similarities.

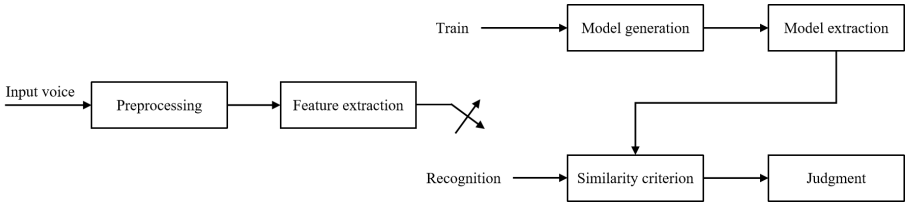

Fig. 1. Block diagram of voice quality recognition system.

4 Acoustic Scene Classification Algorithm[5]

The sound carries a wealth of information about everyday environmental and physical events. Human inherent abilities enable us to perceive our surroundings through sound, such as bustling streets or quiet libraries. Additionally, humans can infer the environment and anticipate events by identifying individual sound sources, such as the sound of a car braking or soothing lullabies. With the widespread use of sound capturing devices like smartphones, an increasing amount of audio is being recorded. The development of signal processing and artificial intelligence technologies allows machines to automatically perceive and extract useful information from acoustic environments, enabling intelligent classification through multidimensional analysis.

Acoustic Scene Classification refers to the use of signal processing and machine learning algorithms to perceive input audio signals and identify the environment in which the audio stream was produced, and label it with semantic tags[6] Ballas[7], using experimental psychology research, found that the recognition speed and accuracy of acoustic events are related to the acoustic properties of the stimuli, such as the frequency of occurrence, the presence of certain physical cues, or the influence of prior sound knowledge. Acoustic scene signal processing involves multiple interdisciplinary fields including digital signal processing, acoustics, auditory psychology, and artificial intelligence, making it a highly comprehensive and cross-disciplinary subject. Research

on acoustic scene classification holds significant practical value. By recording audio in real-time through microphones, identifying acoustic environments, automatically assigning corresponding metadata to audio files, and then rendering acoustic scenes, it can serve as a necessary supplement to videos, thereby enhancing the immersion of VR/AR experiences[8].

The digital audio archives are growing rapidly, containing a vast array of speech, music, animal sounds, urban environmental sounds, and more. Currently, the utilization rate of these audio archives is much lower than that of text and image archives. By using machine learning methods to classify the acoustic scenes within these archives, a wealth of valuable information can be extracted. Acoustic scenes, in comparison to speech and music, lack sustained and distinct harmonic components, encompassing a wider range of acoustic events and a diverse set of signal features. Therefore, for machine listening systems, precise classification of environments is currently unachievable. Figure 2 illustrates the process of extracting acoustic event likelihood features, while Figure 3 depicts the structure of the audio dataset. It is worth noting that the audio dataset structure consists of sound data from common scenarios.

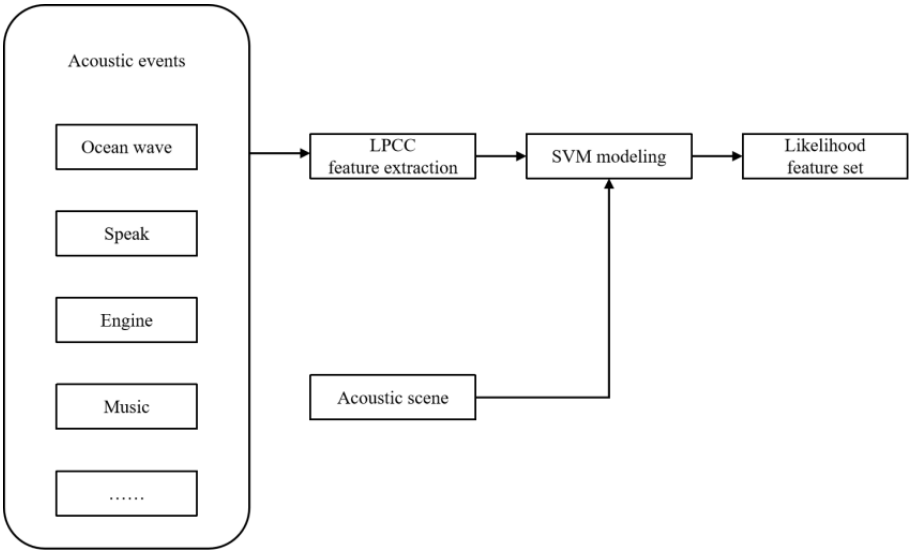

Fig. 2. Extraction process of acoustic event likelihood features.

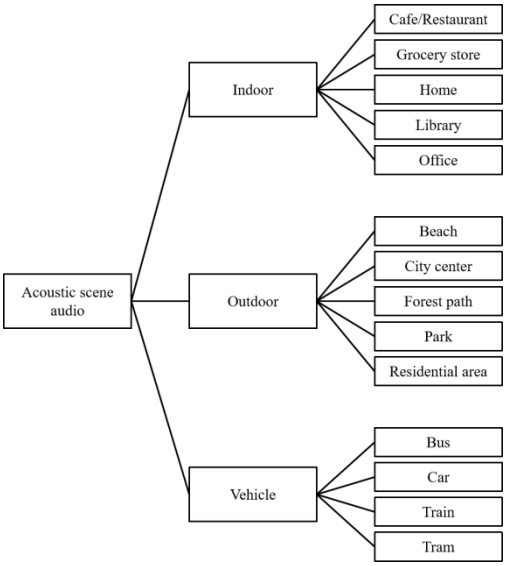

Fig. **3.** Structure diagram of audio data set.

5 Fuzzy Clustering Scoring

In fuzzy clustering, the matrix formed by the rating results vectors of all factors affecting the quality of broadcasting hosts is referred to as the domain. It is utilized through the formula:

$$r_{ij} = \begin{cases} 1(i=j) \\ 1 - C\sum_{k=1}^{n}\left|X_{ik} - X_{jk}\right|(i \neq j) \end{cases} \tag{2}$$

The value of C follows the principle of $0 \leq C \leq 1$, where C represents the scores given by experts to the broadcasting hosts for the primary indicator k. Establish the fuzzy relation matrix R on the domain based on the above formula. We can then determine which methods are most effective for broadcasting hosts and which methods have similar levels of contribution.

6 Conclusion

The application of artificial intelligence (AI) technology has significantly enriched audiences' emotional experiences through visual and auditory enjoyment. This, however, has led to increased demands for hosting and communication. The fundamental functions of hosting communication, including transmission, opinion guidance, public opinion shaping, and cultural guidance, have been further strengthened. Host communication subjects must effectively manage the relationship between intelligent machines and humans, leveraging the role of AI in gathering information and analyzing data to reflect

diverse audience opinions. They must also value verbal language communication to exchange ideas and convey emotions, overcoming algorithmic information push and expanding audience information acceptance. Furthermore, they should focus on the role of sound in cultural heritage and identity, enhancing awareness of using standard language and demonstrating information authority with appealing voices in multimedia integration environments.

Moreover, research on acoustic scene classification needs to draw on cognitive science to understand the human brain's mechanism in distinguishing acoustic scenes and optimize classification algorithms. Exploration of large-scale AI models can drive progress in speech signal processing, enabling complex feature learning, transfer learning, and multimodal task learning.

In conclusion, voice simulation, acoustic scene recognition and classification technology still have promising development prospects in the future. Solving these issues will propel the broadcasting and hosting industry to higher levels. Additionally, the following measures can enhance the effectiveness of artificial intelligence in serving production activities: 1. Improve the algorithm recommendation mechanism to ensure that users can access diverse content and avoid information cocoons. 2. Strengthen the cooperation training between human and AI, so that the host can better use AI technology to improve work efficiency.

References

1. Wang, Y., "Research on hosting and propagation in the field of artificial intelligence and algorithms," Research on Host Communication in China (02), 123-133 (2019).
2. Yang, L., Zeng, J., Dong, G., "Simulation of Speech Enhancement Algorithm Combined with High-Efficiency Channel Attention Mechanism," Computer Simulation 40(03), 258-262+535 (2023).
3. Tu, Z., "Objective evaluation of speech timbre based on auto encoder," Journal of Communication University of China(Science and Technology) 24(04), 8-13 (2017).
4. Gu, Z., Bianba, W., Qi, J., "Tibetan speech emotion recognition based on multi-feature fusion," Modern Electronics Technique 46(21), 129-133 (2023).
5. Pang, X., Ge, F., Li, Y., "Capturing Soundscapes: Exploration and Frontiers of Acoustic Scene Classification in the Digital Era," Computer Engineering 2024, 1-22 (2024).
6. Barchiesi, D., Giannoulis, D., Stowell, D., et al., "Acoustic Scene Classification: Classifying environments from the sounds they produce," IEEE Signal Process. Mag. 32(3), 16-34 (2015).
7. Ballas, James A., "Common factors in the identification of an assortment of brief everyday sounds," J EXP PSYCHOL HUMAN 19(2), 250-267 (1993).
8. Zeng, Z., Pantic, M., Roisman, G I., et al., "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," IEEE Trans. Pattern Anal. Mach. Intell. 31(1), 39-58 (2009).