

Exploring Advancements in Speech Processing

Devansh Handa

Department of Computer Science
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham,
Bengaluru, India.

BL.EN.U4AIE21041@bl.students.amrita.edu

Kindi Krishna Nikhil

Department of Computer Science
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham,
Bengaluru, India.

BL.EN.U4AIE21068@bl.students.amrita.edu

Vaan Amuthu Elango

Department of Computer Science
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham,
Bengaluru, India.

e_vaanamuthu@blr.amrita.edu

Abstract – The purpose of this work is an inventory of the latest discoveries in speech technology represented in twelve bright works. They span from emotional voice conversion to speech analytics and language learning and so on and these studies represent the multidirectional nature of modern experiments. This review studies interconnections amongst domains of speech processing synthesizing key insights which leads to a multi-faceted view of the obstacles, inventions and future projections of speech processing.

Keywords: *Speech processing, emotional voice conversion, speech analytics, language learning, deep learning, machine learning, signal processing, voice conversion techniques, natural language processing, privacy protection, data security, human-machine interaction, computational intelligence.*

I. INTRODUCTION

In speech processing field there are rapid technological development with the key contributions being evident in the intersection of signal processing, machine learning and linguistics. Behind the cutting-edge development, what is now visible is beyond the conception only, for the awareness of the intricate relationship between human language and logical computation now strikes hard upon us. Immerse in these dynamic shifting grounds, emotional voice conversion becomes an interesting research field [1]. By revealing of the subtleties that speech signals contain in the area of emotional expression, researchers seek to match machine cognition with human perception of emotion thus ensuring more caring human-machine interaction.

Also, the strategic integration of speech analytics into diverse domains has completely changed how we draw inferences from audio data streams [2]. From tracking information streams live to empowering seamless customer service, the introduction of automatic keywords detection systems has brought data driven decision making to a whole new level. In addition, due to the demand for efficient language learning approaches, scholars began to examine deep feature-based clustering algorithms [3]. These methods which employ the potential of machine learning provide personalized feedback and correction mechanisms tailored to suit the distinct learning needs and preference of the individual.

Simultaneously, the goal of ensuring the protection of the privacy and security of voice data has given rise to voice de-identification techniques [4]. With voice data growing ever omnipresent in our digital arena, the confidentiality and integrity of sensitive information turns out to be one of the top items on the agenda. Researchers aim to find a middle ground through innovative voice conversion and anonymization methodologies utility and privacy thereby earning trust and an openness in voice-operated systems [5].

Recent studies have been directed into speech and prosodic error annotation and detection, for instance, see [6], [7]. This work may thus lead to more accurate language learning methods accompanied with speech enhancement methods. Likewise, recent advancements in voice conversion methods have brought about expressive voice conversion architectures [8], [9], providing speaker-dependent and high-fidelity synthesis opportunities. Although end-to-end frameworks of speech waveform re-construction and clean content extraction have greatly enhanced the quality and robustness of the speech conversions [10], [11]. Thus, the research of anonymization methods which is based on voice conversion has dealt with the privacy issues formatted in protected speaker characteristics [12].

This paper attempts to offer a synthesis of such diverse views of speech processing beginning from the perception of speech to the advanced applications, focusing on the milestones and the unanswered questions thus far in the evolving field.

REFERENCES

- [1] Shah, N., Singh, M., Takahashi, N., & Onoe, N. (2023). Nonparallel Emotional Voice Conversion for Unseen Speaker-Emotion Pairs Using Dual Domain Adversarial Network & Virtual Domain Pairing.
- [2] Farkhadov, M., Smirnov, V., & Eliseev, A. (2017). Application of Speech Analytics in Information Space Monitoring Systems.
- [3] Sun, H., Wang, D., Li, L., Chen, C., & Zheng, T. F. (2023). Random Cycle Loss and Its Application to Voice Conversion.
- [4] Li, Y. A., Han, C., & Mesgarani, N. (2023). STYLETTS-VC: One-Shot Voice Conversion by Knowledge Transfer from Style-Based TTS Models.
- [5] Scheidt, S., & Chung, Q. B. (2019). Making a Case for Speech Analytics to Improve Customer Service Quality: Vision, Implementation, and Evaluation.
- [6] Nazir, F., Majeed, M. N., Ghazanfar, M. A., & Maqsood, M. (2023). A Computer-Aided Speech Analytics Approach for Pronunciation Feedback Using Deep Feature Clustering.
- [7] Hussain, S., Neekhar, P., Huang, J., Li, J., & Ginsburg, B. (2023). ACE-VC: Adaptive and Controllable Voice Conversion Using Explicitly Disentangled Self-Supervised Speech Representations.
- [8] Ning, Z., Xie, Q., Zhu, P., Wang, Z., Xue, L., Yao, J., Bi, M., & Xie, L. (2023). EXPRESSIVE-VC: Highly Expressive Voice Conversion with Attention Fusion of Bottleneck and Perturbation Features.
- [9] Li, J., Tu, W., & Xiao, L. (2023). FREEVC: Towards High-Quality Text-Free One-Shot Voice Conversion.
- [10] Sisman, B., Yamagishi, J., King, S., & Li, H. (2021). An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning.
- [11] Srivastava, B. M. L., Vauquier, N., Sahidullah, M., Bellet, A., Tommasi, M., & Vincent, E. (2020). Evaluating Voice Conversion-Based Privacy Protection Against Informed Attackers.
- [12] Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, D. L., & Novak, T. P. (2020). Voice Analytics in Business Research: Conceptual Foundations, Acoustic Feature Extraction, and Applications.