World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:18, No:07, 2024

Italian Speech Vowels Landmark Detection through the Legacy Tool 'xkl' with Integration of Combined CNNs and RNNs

Authors : Kaleem Kashif, Tayyaba Anam, Yizhi Wu

Abstract : This paper introduces a methodology for advancing Italian speech vowels landmark detection within the distinctive feature-based speech recognition domain. Leveraging the legacy tool 'xkl' by integrating combined convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the study presents a comprehensive enhancement to the 'xkl' legacy software. This integration incorporates re-assigned spectrogram methodologies, enabling meticulous acoustic analysis. Simultaneously, our proposed model, integrating combined CNNs and RNNs, demonstrates unprecedented precision and robustness in landmark detection. The augmentation of re-assigned spectrogram fusion within the 'xkl' software signifies a meticulous advancement, particularly enhancing precision related to vowel formant estimation. This augmentation catalyzes unparalleled accuracy in landmark detection, resulting in a substantial performance leap compared to conventional methods. The proposed model emerges as a state-of-the-art solution in the distinctive feature-based speech recognition systems domain. In the realm of deep learning, a synergistic integration of combined CNNs and RNNs is introduced, endowed with specialized temporal embeddings, harnessing self-attention mechanisms, and positional embeddings. The proposed model allows it to excel in capturing intricate dependencies within Italian speech vowels, rendering it highly adaptable and sophisticated in the distinctive feature domain. Furthermore, our advanced temporal modeling approach employs Bayesian temporal encoding, refining the measurement of inter-landmark intervals. Comparative analysis against state-of-the-art models reveals a substantial improvement in accuracy, highlighting the robustness and efficacy of the proposed methodology. Upon rigorous testing on a database (LaMIT) speech recorded in a silent room by four Italian native speakers, the landmark detector demonstrates exceptional performance, achieving a 95% true detection rate and a 10% false detection rate. A majority of missed landmarks were observed in proximity to reduced vowels. These promising results underscore the robust identifiability of landmarks within the speech waveform, establishing the feasibility of employing a landmark detector as a front end in a speech recognition system. The synergistic integration of re-assigned spectrogram fusion, CNNs, RNNs, and Bayesian temporal encoding not only signifies a significant advancement in Italian speech vowels landmark detection but also positions the proposed model as a leader in the field. The model offers distinct advantages, including unparalleled accuracy, adaptability, and sophistication, marking a milestone in the intersection of deep learning and distinctive feature-based speech recognition. This work contributes to the broader scientific community by presenting a methodologically rigorous framework for enhancing landmark detection accuracy in Italian speech vowels. The integration of cutting-edge techniques establishes a foundation for future advancements in speech signal processing, emphasizing the potential of the proposed model in practical applications across various domains requiring robust speech recognition systems.

Keywords : landmark detection, acoustic analysis, convolutional neural network, recurrent neural network
Conference Title : ICLAL 2024 : International Conference on Languages and Applied Linguistics
Conference Location : Ottawa, Canada
Conference Dates : July 11-12, 2024