

(12) PATENT APPLICATION PUBLICATION

(21) Application No.202311047010 A

(19) INDIA

(22) Date of filing of Application :12/07/2023

(43) Publication Date : 04/08/2023

(54) Title of the invention : SYSTEM AND METHOD FOR VISUAL SPEECH ENACTMENT USING A NEURAL NETWORK

(51) International classification :G05B 130200, G06N 030400, G06N 030630, G06N 030800, G06T 050000
(86) International Application No :NA
Filing Date :NA
(87) International Publication No : NA
(61) Patent of Addition to Application Number :NA
Filing Date :NA
(62) Divisional to Application Number :NA
Filing Date :NA

(71)Name of Applicant :

1)Chitkara University

Address of Applicant :Chitkara University, Chandigarh-Patiala National Highway, Village Jhansla, Rajpura, Punjab - 140401, India. Patiala -----

2)Bluest Mettle Solutions Private Limited

Name of Applicant : NA

Address of Applicant : NA

(72)Name of Inventor :

1)MISHRA, Saket

Address of Applicant :ODC-4, Panchshil Tech Park, inside Courtyard by Marriott premises, Hinjewadi Phase - 1, Pune - 411057, Maharashtra, India. Pune -----

2)PANDEY, Sakshi

Address of Applicant :ODC-4, Panchshil Tech Park, inside Courtyard by Marriott premises, Hinjewadi Phase - 1, Pune - 411057, Maharashtra, India. Pune -----

3)PANDA, Surya Narayan

Address of Applicant :Chitkara University, Chandigarh-Patiala National Highway, Village Jhansla, Rajpura, Punjab - 140401, India. Patiala -----

(57) Abstract :

The present invention discloses a system (100) and a method (200) for visual speech enactment using a neural network. The system includes a processor (102) to train neural network model to distinguish speech of a visible speaker from background noise using video and corresponding distracting soundtrack, add artificial background noise movies generated from the target speaker's voice, utilize deep neural network techniques for voice augmentation, implement a convolutional neural network model to generate clear spectrograms of improved voice based on lip frames and speech spectrograms. Additionally, the processor (102) may be configured to transform waveform signals using Short-Time Fourier Transform (STFT) and applying mel-scale filtering, and creating spectrograms for speech segments that enable effective enactment of visual speech by separating speech from noise and improving clarity and quality using neural network-based techniques.

No. of Pages : 24 No. of Claims : 10