CLAIMS

1. A server apparatus for processing data, the server apparatus comprises a processor, the processor configured to:

extract a voice activity detection (VAD) module from a Web Real-Time Communication (WebRTC) module, the WebRTC module configured to obtain real-time data input through a web browser;

transplant the VAD module to a Media Resource Control Protocol (MRCP) module, the VAD module configured to perform real-time processing of a plurality of data packets of the real-time data input and return an output indicating either a presence or an absence of a speech activity for each of the plurality of data packets and the MRCP module configured to decide whether to continue processing a subsequent data packet of the plurality of data packets based on the output;

obtain, from the VAD module, an input data stream over a time period, the input data stream comprising audio data;

identify, from the audio data, using a classification algorithm, at least one voice data packet and/or at least one non-voice data packet;

calculate a first consecutive cumulative number of the at least one non-voice data packet over the time period;

calculate a second consecutive cumulative number of the at least one voice data packet over the time period; and

compare the first consecutive cumulative number with a first threshold;

wherein in a positive determination that the first consecutive cumulative number is greater than or equal to the first threshold, output a status indicative of a silent event.

2. The server apparatus of claim 1, wherein the processor is further configured to compare the second consecutive cumulative number with a second threshold, and wherein in a positive determination that the second consecutive cumulative number is greater than or equal to the second threshold, the processor is configured to output a status indicative of a speech start event.

3. The server apparatus of claim 1, wherein the input data stream comprises a plurality of frames, wherein each of the plurality of frames is associated with a single package data, and wherein the classification algorithm comprises a single package data detection and analysis

module, the single package data detection and analysis module configured to perform one or more of the following: a spectral sub-band division, a Gaussian Mixture Model (GMM), a hypothesis testing, and a maximum likelihood estimation, and wherein the identification of the at least one voice data packet and/or at least one non-voice data packet is based on a ratio of noise probability to signal probability.

4. The server apparatus of claim 2 or 3, wherein the processor is further configured to provide interruption control of the VAD module based on the status indicative of the speech start event or the status indicative of the silent event.

5. The server apparatus of claim 4, wherein the interruption control comprises switching the VAD module between a plurality of states.

6. The server apparatus of claim 5, wherein the plurality of states comprise an activity state, an inactivity state, an activity transition state, and an inactivity transition state.

7. The server apparatus of any one of the preceding claims, wherein where the status is indicative of the silent event, the processor is configured to perform automatic speech recognition (ASR) to further determine whether an output of the ASR is indicative of non-speech.

8. The server apparatus of claim 2, wherein where the status is indicative of the speech start event, the processor is configured to determine whether the VAD module is in an initial round of dialogue.

9. The server apparatus of claim 8, wherein in a positive determination that the VAD module is in the initial round of dialogue, the processor is configured to provide a non-interruptible output of the initial round of dialogue.

10. The server apparatus of claim 8, wherein in a positive determination that the VAD module is not in the initial round of dialogue, the processor is configured to obtain a non-interruptible time parameter.

11. The server apparatus of claim 10, wherein the processor is configured to compare the non-interruptible time parameter with a time node parameter of the speech start event, and wherein where the time node parameter exceeds the non-interruptible time parameter, the processor is configured to provide interruption control of the VAD module based on the status indicative of the speech start event.

12. The server apparatus of claim 4, wherein where the interruption control is provided by a switch, and wherein the switch comprises a softphone and a soft switch to provide product drivers for voice and chat.

13. An automated outbound calling system comprising,

an extraction module, the extraction module configured to extract a voice activity detection (VAD) module from a Web Real-Time Communication (WebRTC) module, and transplant the VAD module to a Media Resource Control Protocol (MRCP) module, the WebRTC module configured to obtain real-time data input through a web browser, the VAD module configured to perform real-time processing of a plurality of data packets of the real-time data input and return an output indicating either a presence or an absence of a speech activity for each of the plurality of data packets and the MRCP module configured to decide whether to continue processing a subsequent data packet of the plurality of data packets based on the output;

the VAD module, the VAD module operable to receive an input data stream over a time period, the input data stream comprising audio data;

an audio data processing module configured to:

obtain, from the voice activity detection (VAD) module, the audio data;

identify, from the audio data, using a classification algorithm, at least one voice data packet and/or at least one non-voice data packet;

calculate a first consecutive cumulative number of the at least one non-voice data packet over the time period; and

calculate a second consecutive cumulative number of the at least one voice data packet over the time period; and

compare the first consecutive cumulative number with a first threshold;

wherein in a positive determination that the first consecutive cumulative number is greater than or equal to the first threshold, output a status indicative of a silent event.

14. The automated outbound calling system of claim 13, wherein the audio data processing module is configured to compare the second consecutive cumulative number with a second threshold, and wherein in a positive determination that the second consecutive cumulative number is greater than or equal to the second threshold, output a status indicative of a speech start event.

15. The automated outbound calling system of claim 14, wherein, where the status is indicative of the speech start event, the audio data processing module is configured to determine whether the VAD module is in an initial round of dialogue; and in a positive determination that the VAD module is not in the initial round of dialogue, the audio data processing module is configured to obtain a non-interruptible time parameter.

16. The automated outbound calling system of claim 15, wherein, the audio data processing module is configured to compare the non-interruptible time parameter with a time node parameter of the speech start event, and the automated outbound calling system further comprises an interruption module, the interruption module configured to, where the time node parameter exceeds the non-interruptible time parameter, provide interruption control of the VAD module based on the status indicative of the speech start event.

17. The automated outbound calling system of claim 16, wherein the interruption module is configured to switch the VAD module between a plurality of states, and wherein the plurality of states comprise an activity state, an inactivity state, an activity transition state, and an inactivity transition state.

18. A method for processing data using a processor, comprising the steps of:

extracting a voice activity detection (VAD) module from a Web Real-Time Communication (WebRTC) module, the WebRTC module configured to obtain real-time data input through a web browser;

transplanting the VAD module to a Media Resource Control Protocol (MRCP) module, the VAD module configured to perform real-time processing of a plurality of data packets of the real-time data input and return an output indicating either a presence or an absence of a speech activity for each of the plurality of data packets and the MRCP module configured to

decide whether to continue processing a subsequent data packet of the plurality of data packets based on the output;

obtaining, from a voice activity detection (VAD) module, an input data stream over a time period, the input data stream comprising audio data;

identifying, from the audio data, using a classification algorithm, at least one voice data packet and/or at least one non-voice data packet;

calculating a first consecutive cumulative number of the at least one non-voice data packet over the time period;

calculating a second consecutive cumulative number of the at least one voice data packet over the time period; and

comparing the first consecutive cumulative number with a first threshold;

wherein in a positive determination that the first consecutive cumulative number is greater than or equal to the first threshold, output a status indicative of a silent event.

19. The method of claim 18, further comprises comparing the second consecutive cumulative number with a second threshold and wherein in a positive determination that the second consecutive cumulative number is greater than or equal to the second threshold, output a status indicative of a speech start event.

20. A non-transitory computer-readable medium storing computer executable code comprising instructions for processing data according to the method of claim 18 or 19.