

SERVER APPARATUS, SYSTEM, AND METHOD FOR DETERMINING ONE OR MORE QUALITY RELATED PARAMETER OF A SPEECH DATA PACKET

TECHNICAL FIELD

[0001] Various aspects of this disclosure relate to a server apparatus, system, and method for determining one or more quality related parameters of a speech data packet.

BACKGROUND

[0002] The following discussion of the background is intended to facilitate an understanding of the present disclosure only. It should be appreciated that the discussion is not an acknowledgment or admission that any of the material referred to was published, known, or is part of the common general knowledge of the person skilled in the art in any jurisdiction as of the priority date of the disclosure.

[0003] Currently, manual spot checking remains the mainstream method for speech or voice quality inspection. However, manual spot checking often suffers from inaccuracies when the sample size is small, and may become burdensome and inefficient, leading to significant delays, when the sample size is large. Intelligent quality inspection occupies a small portion of the market. The current intelligent quality inspection technology primarily utilizes the powerful computing capabilities of computers to comprehensively cover voice for quality inspection, but it is generally only applicable to specific scenarios and may lack versatility.

[0004] Accordingly, the disclosure seeks to provide an improved server apparatus, system, and method for determining a parameter of voice packets, such as a quality related parameter.

SUMMARY

[0005] The disclosure seeks to provide a system, server apparatus, and method for processing a speech quality inspection analysis method and system based on ASR, which solves the technical problem of weak versatility in existing intelligent quality inspection technologies due to their applicability only to specific scenarios. By customizing quality inspection rules and scoring rules based on work scenarios to construct a first quality inspection model which is an initial model, and then configuring the initial model with a task list to obtain a second quality inspection model. The ASR speech recognition module is used to convert speech into text information, which is analyzed by the second quality inspection model to

obtain scoring results. The scoring results are then reviewed and revised manually to obtain the final scoring results, and a visual report is generated based on the final scoring results. The customizable quality inspection rules and scoring rules, as well as the model configuration based on the task list, enhance the applicability of the model; while manual review increases the fault tolerance and accuracy of the model. This achieves the technical effect of obtaining a highly versatile intelligent speech quality inspection solution.

[0006] The unified assessment standards among different quality inspectors through the aforementioned speech quality inspection rules reduce human judgment, eliminate subjective influences, ensure objectivity and fairness, and also save labor costs. The intelligent quality inspection of the aforementioned second speech quality inspection model addresses the limitations of quality inspectors' own business capabilities, enhances the effectiveness of quality inspector assessments, and eliminates irregularities in the quality inspection process due to external factors. Intelligent quality inspection can overcome the limitations of quality inspection resources, providing seamless coverage for every customer service call with 100% full quality inspection, effectively avoiding biased sampling. Due to the high efficiency of quality inspection, it can resolve the lag in traditional customer service quality inspection, enabling real-time monitoring of customer service calls, timely detection and early warning, so that supervisors can promptly identify and resolve issues.

[0007] By adopting the technical solution of constructing a first speech quality inspection model based on speech quality inspection rules and quality inspection scoring rules; configuring the first speech quality inspection model according to a task list to obtain a first speech second quality inspection model; converting first speech information awaiting quality inspection into first text information awaiting quality inspection through an ASR speech recognition module; inputting the first text information awaiting quality inspection into the first speech second quality inspection model to obtain a first quality inspection result, wherein the first quality inspection result includes a first quality inspection score; manually reviewing the first quality inspection score through a review instruction to obtain a second quality inspection score; and generating a first quality inspection report based on the second quality inspection score, the technical effect of obtaining a highly versatile intelligent speech quality inspection scheme is achieved. This is done by constructing an initial quality inspection model based on customized quality inspection rules and scoring rules according to work scenarios, and then configuring the initial model with a task list to obtain a second quality inspection model. The ASR speech recognition module is used to convert speech into text information, which is

analyzed by the second quality inspection model to obtain scoring results. The scoring results are then manually reviewed and revised to obtain the final scoring results, and a visual report is generated based on the final scoring results. The customizable quality inspection rules and scoring rules, as well as the model configuration based on the task list, enhance the applicability of the model; while manual review increases the fault tolerance and accuracy of the model.

[0008] According to an aspect of the present disclosure there is provided a server apparatus, the server apparatus comprising a processor, the processor configured to: construct a first speech quality inspection model based on a set of speech quality inspection rules defining one or more criteria used to assess quality of speech and a set of quality inspection scoring rules defining evaluation of speech interactions based on the one or more criteria; configure the first speech quality inspection model based on a task list to obtain a second speech quality inspection model, the task list comprising a set of operations or steps to be performed during a speech quality inspection; obtain a data packet, the data packet comprising speech or voice converted by an automatic speech recognition (ASR) module into text data, into the second speech quality inspection model; and obtain a first quality inspection score based on an output of the second speech quality inspection mode.

[0009] In some embodiments, the processor is configured to obtain a second quality inspection score based on a review of the first quality inspection score.

[0010] In some embodiments, the processor is configured to generate a quality inspection report based on the second quality inspection score.

[0011] In some embodiments, each task in the task list comprises a plurality of tags, each tag associated with one or more of the following: a time node information for each task to be completed, a data volume information for each task to be completed, a quality inspection classification information for each task to be completed.

[0012] In some embodiments, the processor is configured to generate a task allocation information based on the task list, and wherein the task allocation information comprises quality inspection time information and the quality inspection classification information.

[0013] In some embodiments, the processor is configured to obtain a first configuration information based on the quality inspection time information, and/or obtain a second configuration information based on the quality inspection classification information.

[0014] In some embodiments, the second speech quality inspection model is generated based on the first configuration information and the second configuration information.

[0015] In some embodiments, the processor is configured to obtain a modification instruction based on the review of the first quality inspection score, and wherein the modification instruction includes a subtraction instruction and/or an addition instruction, and wherein the processor is further configured to modify the first quality inspection score using the subtraction instruction and/or the addition instruction to obtain the second quality inspection score.

[0016] In some embodiments, the set of speech quality inspection rules comprises one or more manual speech quality inspection rules, and wherein the processor is configured to: convert the one or more manual speech quality inspection rules into a text matching rule set, where the text matching rule set includes word-based rules and phrase-based rules; and generate the speech quality inspection rules using the word-based rules and the phrase-based rules.

[0017] In some embodiments, the processor is further configured to train the first or second speech quality inspection model.

[0018] In some embodiments, the processor is configured to obtain a training dataset, and wherein the training dataset comprises a historical dataset processed by an independent speech quality inspection module, the historical dataset comprising a recording text dataset and a customer feedback information dataset.

[0019] In some embodiments, the processor is configured to obtain a data volume value based on the recording text dataset and the customer feedback information dataset, and obtain a preset data volume value, and determine whether the data volume value is less than the preset data volume value.

[0020] In some embodiments, in a positive determination that the data volume value is less than the preset data volume value, the processor is configured to: obtain one or more manual quality inspection scoring rules.

[0021] In some embodiments, in a negative determination that the data volume value is less than the preset data volume value, the processor is configured to: obtain an automatic generation instruction for the speech quality inspection rules; preprocess the recording text dataset based on the automatic generation instruction for the set of speech quality inspection rules to obtain a recording text information, wherein the second recording text dataset includes customer recording text and customer service recording text.

[0022] In some embodiments, the processor is configured to encode the customer recording text and the customer service recording text using an encoding instruction to obtain an encoding result.

[0023] In some embodiments, the processor is further configured to generate a classification label based on a business type of the second recording text dataset, and identify the encoding result based on the classification label to generate the quality inspection scoring rules.

[0024] According to another aspect of the present disclosure there is provided a system for speech quality inspection analysis, the system comprises an automatic speech recognition module; the server apparatus of any one of the preceding claims, the server apparatus arranged in data communication with the automatic speech recognition module to receive a data packet, the data packet comprising speech or voice data converted by the automatic speech recognition module into text data; and wherein the server apparatus is configured to display visual information, the visual information comprising the first quality inspection score.

[0025] In some embodiments, the visual information further comprises voice text information, and metadata.

[0026] According to another aspect of the present disclosure there is provided a method for determining a parameter of speech data packets, the method comprises: constructing a first speech quality inspection model based on a set of speech quality inspection rules defining one or more criteria used to assess the quality of speech and a set of scoring rules defining evaluation of speech interactions based on the one or more criteria; configuring the first speech quality inspection model based on a task list to obtain a second speech quality inspection model, the task list comprising a set of operations or steps to be performed during a speech quality inspection; obtaining a data packet, the data packet comprising speech or voice converted by an automatic speech recognition (ASR) module into text data, inputting the data packet into the second speech quality inspection model; and obtaining a first quality inspection score based on an output of the second speech quality inspection model.

[0027] According to another aspect of the present disclosure there is provided a non-transitory computer-readable medium storing computer executable code comprising instructions for processing data according to any one of the described methods.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] The disclosure will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

- FIG. 1A is a schematic block diagram of a server apparatus for determining a parameter of a speech data packet, such as one or more quality inspection scores of the speech data packet.
- FIG. 1B is a schematic block diagram comprising the server apparatus of FIG. 1A, implemented as a system for determining one or more quality inspection scores of the speech data packet.
- FIG. 2 is a flow chart depicting an ASR-based speech quality inspection and analysis method according to an embodiment.
- FIG. 3 is a flow chart illustrating a specific embodiment for configuring the first speech quality inspection model based on a task list to obtain the second speech quality inspection model.
- FIG. 4 is a flow chart illustrating a specific embodiment of a method for constructing the first speech quality inspection model based on the set of speech quality inspection rules and the set of scoring rules.
- FIG. 5 is a flow chart illustrating a method for generating scoring rules based on historical data according to an embodiment.
- FIG. 6 is a flow chart illustrating a method for obtaining or generating a first quality inspection score.
- FIG. 7 is a flow chart illustrating a method for reviewing the first quality inspection score to obtain the second quality inspection score.
- FIG. 8 is a general flow chart of a method for determining one or more quality inspection scores of the speech data packet according to the present disclosure.

DETAILED DESCRIPTION

[0029] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the disclosure may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the disclosure. Other embodiments may be utilized and structural, and logical

changes may be made without departing from the scope of the disclosure. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0030] Embodiments, which are non-limiting examples, described in the context of one of the enclosure systems, server devices, or methods are analogously valid for the other systems, devices, or methods. Similarly, embodiments described in the context of a system are analogously valid for a device or a method, and vice-versa.

[0031] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0032] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements. Furthermore, as used in the present disclosure and the appended claims, the term “by” may also mean “from”, depending on the context. Furthermore, as used in the present disclosure and the appended claims, the term “if” may also mean “when” or “upon”, depending on the context. Furthermore, as used in the present disclosure and the appended claims, the words “and/or” may refer to and encompass any and all possible combinations of one or more of the associated listed items.

[0033] As used herein, the term “data” may be understood to include information in any suitable analogue or digital form, for example, provided as a file, a portion of a file, a set of files, a signal or stream, a portion of a signal or stream, a set of signals or streams, waveforms, and the like. The term data, however, is not limited to the aforementioned examples and may take various forms and represent any information as understood in the art.

[0034] As used herein, the term “first”, “second”, “third”, “fourth”, “fifth”, etc. are used to distinguish one element/feature from another, and, unless otherwise stated, may not denote order, priority or sequence.

[0035] As used herein, the term “module” refers to, forms part of, or includes an application Specific Integrated Circuit (ASIC); an electronic circuit; a combinational logic circuit; a field programmable gate array (FPGA); a processor (shared, dedicated, or group) that executes code; other suitable hardware components that provide the described functionality; or a combination

of some or all of the above, such as in a system-on-chip. The term module may include memory (shared, dedicated, or group) that stores code executed by the processor. A single module or a combination of modules may be regarded as a device. A processor may include one or more modules. For example, multiple modules described in this disclosure may form a processor.

[0036] As used herein, the term “associate”, “associated”, and “associating” indicate a defined relationship (or cross-reference) between two items.

[0037] As used herein, the term “obtain”, in the context of obtaining data, broadly include pull technology used any time a transfer of data is initiated by a request sent from a client to a server. Push technology, on the other hand, is implemented any time a transfer of information is initiated by a server without waiting on a request from a client. In some embodiments, the term obtain may include receive.

[0038] As used herein, “memory” may be understood as a non-transitory computer-readable medium in which data or information can be stored. References to “memory” included herein may thus be understood as referring to volatile or non-volatile memory, including random access memory (“RAM”), read-only memory (“ROM”), flash memory, solid-state storage, magnetic tape, hard disk drive, optical drive, etc., or any combination thereof. Furthermore, it is appreciated that registers, shift registers, processor registers, data buffers, etc., are also embraced herein by the term memory. It is appreciated that a single component referred to as “memory” or “a memory” may be composed of more than one different type of memory, and thus may refer to a collective component including one or more types of memory. It is readily understood that any single memory component may be separated into multiple collectively equivalent memory components, and vice versa. Furthermore, while memory may be depicted as separate from one or more other components (such as in the drawings), it is understood that memory may be integrated within another component, such as on a common integrated chip.

[0039] As used herein, the term “Automatic speech recognition (ASR)” or ASR module refers to a system that converts spoken language into written text. An ASR system may utilize algorithms and computational models to process audio input, identify phonetic components, and map them to corresponding words or phrases in a given language. ASR systems may be used in applications such as voice assistants, transcription services, voice-controlled systems, and real-time captioning. In some embodiments, the ASR model may include continuous speech recognition (CSR), dysarthric speech recognition (DSR), rule-based, statistical, and hybrid models.

[0040] As used herein, the term “speech quality inspection model” includes any mathematical or computational model used to evaluate and assess the performance, accuracy, and reliability of an ASR system. These speech quality inspection models may be designed to quantify various aspects of speech recognition output, such as word error rates, character error rates, phonetic accuracy, intelligibility, and overall quality. These models may further be used for determining the accuracy, intelligibility, and usability of ASR systems under varying conditions, such as diverse accents, noisy environments, and complex linguistic structures. For example, an error rate (WER) model measures the number of words or phrases incorrectly recognized by the ASR system, a character error rate (CER) model evaluates the number of individual characters that are incorrect in the recognition output, a phonetic accuracy model assesses how accurately the ASR system recognizes phonemes, syllables, and other units of speech sound, an intelligibility score model measures how easily a person can understand the spoken language from the ASR output, a Signal-to-Noise Ratio (SNR) Model: evaluates the ratio of desired signal quality to background noise in audio recordings used for training or testing the ASR system, a perceptual evaluation form (PEF) model assesses a human listener’s perception and subjective evaluation of speech recognition output, such as clarity, intelligibility, and naturalness, a mean opinion score (MOS) model measures a set of scores from multiple listeners to evaluate the overall quality and acceptability of ASR output, a dynamic time warping (DTW) model evaluates the similarity between two speech signals by comparing their temporal patterns and rhythms, a mel-frequency cepstral coefficient (MFCC) model analyzes the acoustic characteristics of speech, such as pitch, tone, and spectral features, a speech quality assessment tool (SQAT) is a comprehensive framework for evaluating various aspects of ASR performance. It is contemplated that the speech quality inspection model may include machine learning-based models, i.e. models employing supervised or unsupervised learning algorithms to predict speech quality, including deep neural networks (DNNs) trained on large datasets to classify or score speech quality, convolutional neural networks (CNNs) used for spatial feature extraction from spectrograms or waveforms, recurrent neural networks (RNNs) and transformers used to capture temporal dependencies in speech signals. It is further contemplated that the speech quality inspection model may include various permutations and/or combinations of the aforementioned models.

[0041] As used herein, the term “speech quality inspection rules” may be used to evaluate the accuracy and performance of a speech recognition system. These rules may include word error rate (WER), which may be a metric used to evaluate ASR systems. It measures the edit

distance between the recognized text and the reference transcript, phoneme error rate (PER), similar to WER, but at the phoneme level, useful for assessing pronunciation accuracy, speaker diarization error rate (DER) used in multi-speaker scenarios to evaluate how well the system identifies and separates different speakers, real-time factor (RTF): Measures the processing speed of the ASR system relative to the duration of the input audio.

[0042] As used herein, the term “quality inspection scoring rules” may include various methods to quantify and interpret the results of the aforementioned inspection rules. Non-limiting examples include WER scoring, confidence score thresholds, RTF scoring, and/or custom scoring systems.

[0043] As used herein, the term “configured to” broadly refers to an arrangement, a design, and/or a program to perform a specific function. It implies a purposeful arrangement or adaptation for achieving the stated functionality. For example, a processor configured to process data may include hardware (e.g. electronic circuitry, chips), and/or software working in tandem to process the data.

[0044] According to an aspect of the disclosure and with reference to FIG. 1, there is provided a server apparatus for processing speech data packets. The server apparatus may be part of a distributed system, the server apparatus arranged or operable to process the speech data packets to determine and analyse speech quality from the speech data packets. The server apparatus may be used to determine speech quality and may form part of a system for speech quality inspection and analysis.

[0045] The server apparatus may comprise a processor and a memory, the processor is capable of being configured to execute instructions stored in the memory to receive a speech data packet. In the embodiment illustrated in FIG. 1A and FIG. 1B, the server apparatus may be a communications server apparatus. The communications server apparatus may be in the form of a server computer 100, the server computer 100 may be a single server as illustrated schematically in FIG. 1A, or have the functionality performed distributed across multiple server components.

[0046] In some embodiments, the server computer 100 may include an automatic speech recognition (ASR) module configured to receive audio data for conversion to text data. In other embodiments, the server computer 100 may be arranged in data communication with an ASR system and/or configured to receive text data from the ASR system.

[0047] In some embodiments, the server computer 100 includes a communication interface 102. The communication interface 102 may be configured to send and receive data, which may

include the speech data packets, the speech data packets may further include audio data packets and/or text (converted from audio) data packets. The communication interface 102 may include a transmitter module and/or a receiver module allowing the server apparatus to communicate over a communications network. The communication interface 102 may include one or more user-interfaces configured to provide users for user control and may include, for example, one or more computing peripheral devices such as display monitors, computer keyboards and the like.

[0048] The server computer 100 may further include a processor in the form of processing unit 104 and a memory 106. The memory 106 may be used by the processing unit 104 to store, for example, the speech data packets, historical quality data associated with similar speech data packets, and/or quality-based parameters such as scores, to be processed.

[0049] As shown in FIG. 1B, the processing unit 104 may be configured to receive, from a terminal device 110, a speech data packet 111 for processing. The speech data packet 111 may be parsed to obtain information or data relating to speech, non-speech. The non-speech information may include background sound or noise.

[0050] The server computer 100 and one or more terminal devices 110 may be connected via a network 120 to form a system 150 for processing the data. The network 120 may be an Internet or Intranet network. In some embodiments, the system 150 may comprise one or more ASR modules 130, the ASR module 130 may send the speech data packet 111 in text format.

[0051] In some embodiments, the server computer 100 may construct a first speech quality inspection model based on a set of speech quality inspection rules 112 and a set of scoring rules 113. The first speech quality inspection model may be regarded as an initial model for subsequent modification. In some embodiments, the first speech quality inspection model may refer to an original quality inspection model constructed by a development end of a speech quality inspection system based on a basic set of data.

[0052] The server computer 100 may further configure the first speech quality inspection model based on a task list 114 to obtain a second speech quality inspection model. The task list 114 may be a structured component used for configuring and organizing the inspection process. In some embodiments, the task list 114 may include a set of operations or steps to be performed during the inspection, characteristics or criteria to be evaluated for each operation, and/or specific instructions for quality control personnel. The second speech quality inspection model may be regarded as a customized or working model based on the task list 114.

[0053] It may be appreciable that the configuration of the first speech quality inspection model according to the task list 114 may facilitate an alignment or unification of assessment standards. In some embodiments, the task allocation information in the task list, such as, but not limited to, quality inspection time information and quality inspection classification information, etc. may be used to provide guidance that the quality inspection activities are carried out around the established standards. One or more tasks may be arranged based on the above information, making the overall quality inspection work orderly and in line with the unified requirements. In some embodiments, the configuration of the first speech quality inspection model may allow for diverse task arrangements, and concurrently the quality inspection standards for these tasks may be aligned, based on the task list, to be consistent with the overall quality assessment standards of the user/enterprise. The speech quality inspection rules 112, the set of scoring rules 113, and the task list 114 may be obtained from one or more databases 1000A, 1000B. It is appreciable that the databases 1000A, 1000B are for illustration purpose. In some embodiments, the speech quality inspection rules 112, the set of scoring rules 113, and the task list 114 may be obtained from different databases. In some embodiments, the speech quality inspection rules 112, the set of scoring rules 113, and the task list 114 may be obtained from a single databases. In some embodiments, the databases may be arranged in distributed configuration.

[0054] The processing unit 104 may then be operable/configured to obtain or receive the speech data packet 111. The speech data packet 111 may comprise converted speech or voice into text data. The processing unit 104 may then derive or determine a first quality inspection score 115 based on an output of the second speech quality inspection model.

[0055] In some embodiments, the processing unit 104 may be configured to obtain a second quality inspection score 116 based on a review of the first quality inspection score 115.

[0056] In some embodiments, the processing unit 104 may be configured to generate a quality inspection report 117 based on the second quality inspection score 116.

[0057] In some embodiments, each task in the task list 114 may comprise a plurality of tags, each tag associated with one or more of the following: a time node information for each task to be completed, a data volume information for each task to be completed, a quality inspection classification information for each task to be completed.

[0058] In some embodiments, the processing unit 104 may be configured to generate a task allocation information based on the task list, and wherein the task allocation information

comprises quality inspection time information and the quality inspection classification information.

[0059] In some embodiments, the processing unit 104 may be configured to obtain a first configuration information based on the quality inspection time information, and/or obtain a second configuration information based on the quality inspection classification information.

[0060] In some embodiments, a second speech quality inspection model is generated based on the first configuration information and the second configuration information.

[0061] In some embodiments, the processor is configured to obtain a modification instruction based on the review of the first quality inspection score, and wherein the modification instruction includes a subtraction instruction and/or an addition instruction, and wherein the processor is further configured to modify the first quality inspection score using the subtraction instruction and/or the addition instruction to obtain the second quality inspection score.

[0062] In some embodiments, the set of speech quality inspection rules comprises one or more manual speech quality inspection rules, and wherein the processor is configured to: convert the one or more manual speech quality inspection rules into a text matching rule set, where the text matching rule set includes word-based rules and phrase-based rules; and generate the speech quality inspection rules using the word-based rules and the phrase-based rules.

[0063] In some embodiments, the processing unit 104 may be further configured to train the first or second speech quality inspection model.

[0064] In some embodiments, the processing unit 104 may be configured to obtain a training dataset, and wherein the training dataset comprises a historical dataset processed by an independent speech quality inspection module, the historical dataset comprising a recording text dataset and a customer feedback information dataset.

[0065] In some embodiments, the processing unit 104 may be configured to obtain a data volume value based on the recording text dataset and the customer feedback information dataset, and obtain a preset data volume value, and determine whether the data volume value is less than the preset data volume value.

[0066] In some embodiments, in a positive determination that the data volume value is less than the preset data volume value, the processor is configured to: obtain a manual quality inspection scoring rules, and in a negative determination that the data volume value is less than the preset data volume, the processor is configured to: obtain an automatic generation

instruction for quality inspection rules; preprocess the recording text dataset based on the automatic generation instruction for quality inspection rules to obtain a recording text information, wherein the second recording text dataset includes customer recording text and customer service recording text.

[0067] In some embodiments, the processing unit 104 may be configured to encode the customer recording text and the customer service recording text using an encoding instruction to obtain an encoding result.

[0068] In some embodiments, the processing unit 104 may be further configured to generate a classification label based on the business type of the second recording text dataset, and identify the encoding result based on the classification label to generate the quality inspection scoring rules.

[0069] FIG. 2 is a flow chart depicting an ASR-based speech quality inspection and analysis method 200. The method may be implemented as executable software codes stored in one or more non-transitory computer-readable medium of the processing unit 104 and/or server computer 100. The method 200 may be applied to a speech quality inspection system, the system includes an ASR module. The method 200 may comprise the steps of:

[0070] Step S201: constructing a first speech quality inspection model based on the speech quality inspection rules and set of scoring rules.

[0071] The speech quality rule may be customized quality inspection rule used to construct the first speech quality inspection model, which may be regarded as an initial model for subsequent development and fine-tuning to obtain a working model. The first speech quality inspection model may be an intelligent quality inspection model.

[0072] The speech quality inspection rules may include, but is not limited to: the name of the quality inspection rule, adding regular expressions for the corresponding quality inspection rule name, assigning rule scores to the corresponding quality inspection rules, and returning labels for the quality inspection rules that are matched. In some embodiments, through the customized content of the mentioned first speech quality inspection rule, one or more rules can be configured for the speech speed, volume, silence, interruption, and other situations of different users, such as, but not limited to, agents or customers. The mentioned scoring rules refers to the mechanism of assigning scores to different speech quality inspection rules and the point addition and deduction mechanism when the corresponding quality inspection rules are matched, constructing customized rules for the scoring mechanism. As an example, a scoring rule can assign scores to different speech quality inspection rules and calculate a total score

based on the assigned or matched results. The first speech quality inspection model may refer to the original quality inspection model constructed by the development end of the speech quality inspection system based on the basic data provided by the user end, customized according to the mentioned first speech quality inspection rule and the mentioned first quality inspection scoring rule. By allowing customized quality inspection rules for the user end, the applicability and individualization of intelligent quality inspection may be enhanced.

[0073] In some embodiments, the score may be based on a 10 point or 100-point system.

[0074] In step S202, configuring the first speech quality inspection model based on a task list to obtain a second speech quality inspection mode, i.e. a working model.

[0075] After a development end of the speech quality inspection system has completed the construction of the first speech quality inspection model, and the first speech quality inspection model has passed a stability test, the first speech quality inspection model may be sent to the user end of the speech quality inspection system. In some embodiments, the task list may comprise one or more tasks (i.e. task list information) that needs to be processed when the user end of the speech quality inspection system is in use. The second speech quality inspection model may refer to the intelligent quality inspection model used for work, which is obtained after allocating and configuring tasks for the first speech quality inspection model by inputting the task list each time the first speech quality inspection model is used, and correspondingly associating the configured tasks with the corresponding functional modules for easy invocation. One or more users can customize the configuration of the first initial speech model based on the task list to obtain a working model that is more suitable for them, which enhances versatility.

[0076] In some embodiments, "stability" refers to the ability of the first speech quality inspection model to consistently and accurately output expected results when processing speech quality inspection tasks. Stability may be regarded as an indicator for evaluating model performance, directly related to the reliability and effectiveness of the model in practical applications. In some embodiments, in order to achieve preset stability, the following steps may be carried out.

[0077] Model training: The first speech quality inspection model may be trained with a large amount of speech data, enabling it to learn speech characteristics and quality inspection rules in different scenarios.

[0078] Cross-Validation: The first speech quality inspection model may be evaluated using cross-validation methods to ensure that it performs well on different datasets.

[0079] Parameter-tuning: The parameters of the first speech quality inspection model may be optimized to find/determine the optimal parameter combination, thus improving model stability and accuracy.

[0080] Testing and validation: After model training, an independent test dataset may be used to test and validate the first speech quality inspection model, ensuring that it can accurately output quality inspection results in practical applications.

[0081] In step S203, converting, by an ASR module, data packet comprising speech data or information awaiting quality inspection into text data or information awaiting quality inspection through the ASR speech recognition module. In the step S203, the data packet to be inspected refers to the speech information that is input to the second speech quality inspection model. The text data to be inspected may include structured text information that is obtained by converting the speech data or information to be inspected into text data or information and then preferably processing the text information. The text data may then be input through a natural language processing (NLP) module, such as, but not limited to, Chinese word segmentation and text annotation, thus making it recognizable by computers. In some embodiments, the text data may include text information of speech from different user groups, for example, text information from customer service and the text information from customers. In some embodiments, the text data that is converted by the ASR module may involve two people, and the type of vocabulary to be recognized in the usage scenario may be relatively less diverse and has relatively low complexity. Therefore, the ASR module can efficiently and accurately convert the speech data to be inspected into the text data to be inspected.

[0082] In step S204, inputting the text data awaiting quality inspection into the second speech quality inspection model (i.e. the working model), to obtain the first quality inspection result, wherein the first quality inspection result includes a first quality inspection score.

[0083] The first quality inspection result refers to the quality inspection result obtained by inputting the text data to be inspected into the second speech quality inspection model when the conversion of the text data to be inspected is completed and reaches a quality inspection time node in the task list. The first quality inspection score may refer to a scoring result, which may be in the form of a number, recorded as the first quality inspection result after each speech data is inspected by the second speech quality inspection model. By reviewing the first quality inspection score, a user can understand the quality inspection rules and the point adjustments (additions or deductions) for each speech, thereby analyzing the rules that the corresponding

customer service personnel need to improve and maintain, ultimately enhancing customer service quality.

[0084] It is appreciable that even if the quality inspection time node (e.g. time pattern standard) is not met, the voice/speech text information may still be input into the second quality inspection model for scoring, but it may not be processed immediately.

[0085] Step S205: Reviewing the first quality inspection score through a review instruction to obtain a second quality inspection score. The further review may be a manual review.

[0086] FIG. 3 is a flow chart illustrating a method 300 for configuring configuring the first speech quality inspection model. The method 300 may include:

[0087] Step S301: obtaining a task allocation information based on the task list, wherein the task allocation information includes a quality inspection time information and a quality inspection classification information.

[0088] Step S302: obtaining the first configuration information based on the quality inspection time information.

[0089] Step S303: obtaining a second configuration information based on the quality inspection classification information.

[0090] Step S304: configuring the first speech quality inspection model through the first configuration information and the second configuration information to generate the second speech quality inspection model.

[0091] In some embodiments, the task list includes task list information that a user may be required to complete for quality inspection, and each task in the task list may include multiple tags. Each of the multiple tags may include one or more of the following information: time node information for each task to be completed, data volume information for each task to be completed, quality inspection classification information for each task to be completed, etc.

[0092] In some embodiments, the task allocation list may include one or more results obtained by allocating tasks in the task list based on the tag information in the task list, which may be divided into the quality inspection time information and the quality inspection classification information. The first configuration information may include configuring quality inspection time nodes for different tasks in the first speech quality inspection model based on the quality inspection time information. A non-limiting example of the configuration method may include: allocating, based on the time node information for each task to be completed, which is divided into two types: scheduled execution and immediate execution. Immediate execution is preferably performed in sequential order based on time nodes; scheduled execution

is preferably performed in sequential order based on time nodes after reaching the preset schedule. In some embodiments, the second configuration information may include configuring different tasks belonging to different business types to respond to callable working modules in the first speech quality inspection model based on the quality inspection classification information, which improves the efficiency of quality inspection.

[0093] After the first configuration information and the second configuration information are configured, the second speech quality inspection model is obtained. The second speech quality inspection model can then be used for efficient intelligent voice quality inspection.

[0094] FIG. 4 is a flow chart illustrating a specific embodiment of a method 400 for construct the first speech quality inspection model based on the set of speech quality inspection rules and the set of scoring rules.

[0095] Step S401: obtaining one or more manual speech quality inspection rules based on a user.

[0096] Step S402: converting the one or more manual speech quality inspection rules into a text matching rule set, where the text matching rule set includes word-based rules and phrase-based rules.

[0097] Step S403: Generating the speech quality inspection rules using the word-based rules and the phrase-based rules.

[0098] In some embodiments, the user may include user end of the speech quality inspection system, which can be an enterprise, individual, organization, or other entities. The manual speech quality inspection rules may include various standards used by the user to conduct manual speech quality inspections in a traditional manner. The standards may include the ITU-T P.8 standard, the G.168 standard, the ANSI S4.24-1988 standard, and/or the ETSI ETS 300 335 standard. The text matching rule set may be a predefined set of matching rules obtained based on the part-of-speech and meaning of words. The word-based rules and phrase-based rules may be part of the text matching rule set and may include predefined word-based rules, phrase-based rules, and script rules. In some embodiments, the one or more speech quality inspection rules are formed based on the text matching rule set. An example of the working process during quality inspection is as follows: During quality inspection, target words in the text information of speech to be inspected, the text information may include keywords, sensitive words, prohibited words, etc., are combined into rule expressions based on the text matching rule set. These rule expressions can be used to detect dialogue logic, service processes, and other dialogue-related content. The rule expressions are automatically compared

and matched with the word-based rules and the phrase-based rules. In some embodiments, only when a rule expression matches at least one word and at least one phrase, it indicates that the rule expression has matched the quality inspection rules and the matching is successful. Otherwise, it indicates a miss and a failed match. By customizing the speech quality inspection rules based on the information of the user, the applicability of the speech quality inspection system is improved.

[0099] FIG. 5 is a flow chart illustrating an embodiment of a method 500 for generating scoring rules based on the following steps.

[00100] Step S501: obtaining a training dataset, where the training dataset comprises a historical dataset processed by an independent speech quality inspection module, the historical dataset comprising a recording text dataset and a customer feedback information dataset.

[00101] Step S502: obtaining a data volume value based on the recording text dataset and the customer feedback information dataset.

[00102] Step S503: obtaining a preset data volume value and determine whether the data volume value is less than the preset data volume value, for example, by comparison.

[00103] Step S504: in a positive determination that the data volume value is less than the preset data volume value, obtaining one or more manual quality inspection scoring rules from the user.

[00104] In some embodiments, the step S504 may include:

[00105] Step S505: If the data volume value is not less than (i.e. negative determination) the preset data volume value, obtaining an automatic generation instruction for quality inspection rules.

[00106] Step S506: Preprocessing the recording text dataset based on the first automatic generation instruction for quality inspection rules to obtain second recording text dataset. The second recording text dataset may include a customer recording text and a customer service recording text.

[00107] Step S507: Encoding the customer recording text and the customer service recording text using an encoding instruction to obtain an encoding result.

[00108] Step S508: Generating a classification label based on the business type of the second recording text dataset.

[00109] Step S509: Identifying the encoding result based on the classification label to generate the quality inspection scoring rules.

[00110] Specifically, the determination that the data volume value is not less than the preset data volume value may indicate that the data volume is sufficient for fully training and refining one or more intelligent autonomous inductive scoring rules, thereby generating the quality inspection scoring rules. It is contemplated that the specific value for the preset data volume can vary depending on factors such as the complexity of the business scenarios, the diversity of speech patterns, and the desired accuracy of the scoring model. As a non-limiting example, a value of 10,000 data points could be set as data volume. This value assumes a balanced dataset containing both customer and agent speech texts, along with corresponding customer feedback. With 10,000 data points, the system to be trained (i.e. the first and/or second speech quality inspection model) may have a reasonable sample size to identify patterns, classify different scenarios, and assign appropriate scores without overfitting or underfitting the model. In general, a preset data volume value is chosen to be large enough to capture the variability in speech interactions but not excessively large to make the training process computationally infeasible.

[00111] In some embodiments, the automatic generation instruction for quality inspection rules may refer to the generation of a control signal when the data volume value is not less than the preset data volume value. The second recording text dataset may refer to the result obtained by preprocessing the first recording text data set after the speech quality inspection system receives the first automatic generation instruction for quality inspection rules. A non-limiting example of preprocessing methods is using data cleaning methods such as regular expressions to remove special symbols, modal particles, stop words, and other words that are not useful for the algorithm, which can reduce the computational load and improve accuracy.

[00112] In some embodiments, the second recording text dataset may be divided into the customer recording text and the customer service recording text. Moreover, the encoding instruction may include a control signal issued after obtaining the second recording text dataset. The encoding result may refer to the result obtained by encoding the customer recording text and the customer service recording text after receiving the encoding instruction. The encoding result is data that can be recognized and calculated by a computer to represent the information in the customer recording text and the customer service recording text. The classification label may be generated by selecting different evaluation category label information based on the business type of the second recording text dataset. Identifying the encoding result based on the classification label can represent different types of evaluation categories corresponding to different customer recording text and customer service recording text information. When the

corresponding quality inspection rules are matched, the corresponding evaluation category can be invoked for scoring, thereby automatically generating the quality inspection scoring rules. The automatically generated quality inspection scoring rules are based on a large amount of data and are representative and accurate.

[00113] Step S510: obtaining the quality inspection scoring rules based on the manual quality inspection scoring rules.

[00114] In some embodiments, the training dataset may include a collection of speech data that has been quality inspected by the speech quality inspection system at the user end. The recording text dataset may include recorded text data of customer service personnel during their work that has been quality inspected by one or more speech quality inspection systems. The customer feedback information dataset may include feedback information from customers regarding the performance of customer service personnel. The data volume value may include data volume calculated based on the combination of the recording text dataset and the customer feedback dataset. The preset data volume value may provide an indication of a minimum data volume preset to determine whether it is necessary to manually set the quality inspection scoring rules. When the data volume value is less than the preset data volume value, there is an indication that the data is insufficient and cannot fully train and refine the intelligent autonomous inductive scoring rules. Therefore, it is necessary to manually set the quality inspection scoring rules, so the first manual quality inspection scoring rules are retrieved. The second speech quality inspection model may be configured to score the speech text information to be inspected based on the manual quality inspection scoring rules. When the data volume is insufficient and cannot fully train and refine the intelligent autonomous inductive scoring rules, scoring may be performed using the manual quality inspection scoring rules to avoid inaccurate quality inspection results due to insufficient data volume.

[00115] In other words, the preset data volume value may be regarded as a preset criterion used to determine whether manual setting of the quality inspection scoring rules is required. It is a minimum data volume threshold. When the historical data volume processed by the system is less than this threshold, the system considers the data volume insufficient to support the accurate automatic generation of quality inspection scoring rules, thus requiring reliance on the user's manual quality inspection scoring rules.

[00116] As the preset data volume value is set according to an actual system situation and requirements, it may be adjusted depending on system requirements. Assuming that the speech quality inspection system is mainly used to process customer service recordings, and the system

has accumulated a certain amount of historical data, a possible reference value of the preset data volume value may be 5,000 recorded text data entries (or corresponding data volume units such as GB, MB, etc., depending on the actual situation of data storage and processing). This means that if the combined data volume of the system's current recorded text dataset and customer feedback information dataset is less than 5,000 entries, the system will consider the data volume insufficient and need to rely on the user's manual quality inspection scoring rules to generate the quality inspection scoring rules.

[00117] An embodiment of a method 600 for obtaining or generating the quality inspection score may be illustrated with reference to the flowchart illustrated in FIG. 6, which comprises the following steps:

[00118] Step S601: Matching each of the speech quality inspection rules with each speech quality inspection scoring rule to obtain the one or more matching results;

[00119] Step S602: Obtaining the first quality inspection score through the quality inspection scoring rules, wherein the first quality inspection score corresponds one-to-one with the classification label;

[00120] Step S603: Assigning scores to the first voice quality inspection rule based on the first matching result through the first quality inspection score, to obtain the first scoring standard;

[00121] Step S604: Generating the first speech quality inspection model according to the first scoring standard.

[00122] The first matching result refers to the result obtained by matching the first voice quality inspection rule with the first quality inspection scoring rule. Different keywords, phrases, or scripts may correspond to different quality inspection rules and also correspond to different scoring standards. Based on this, the first voice quality inspection rule can be matched with the first quality inspection scoring rule; the first quality inspection score refers to the result of extracting the score information corresponding to different classification labels in the first quality inspection scoring rule; the first scoring standard refers to the result of extracting the different quality inspection rules corresponding to the classification labels in the first voice quality inspection rule, and then assigning scores to the corresponding quality inspection rules based on the first quality inspection score. A non-limiting example of the assignment method is as follows: the voice quality inspection system may have a default base score of 100 points, and when scoring a certain customer service dialogue, adjustments are made based on this 100 points. For example, if a certain dialogue hits the "bad tone" in the first voice quality inspection

rule, which is set to -5 points in the system, and also hits the "appropriate speaking speed" in the first voice quality inspection rule, which is set to 1 point in the system, then the overall score for this dialogue would be $SUM=100+(-5)+(1)=96$, without considering other hit rules. The assignment of this score can be customized by the user according to their different business types; further, the first speech quality inspection model is generated based on the first scoring standard. By matching the first voice quality inspection rule with the first quality inspection scoring rule and assigning scores to the first voice quality inspection rule, a customizable scoring system is constructed, achieving the technical effect of improving the scope of application. In summary, the equation is based on the base score and the rules for adding and subtracting points. The scoring model can assign different scores to dialogues based on different quality inspection rules, thereby comprehensively evaluating the quality of the dialogue. At the same time, the setting of the base score and the rules for adding and subtracting points can be customized according to actual needs, improving the applicability and flexibility of the model.

[00123] In some embodiments, the review process to obtain the second quality inspection score may be based on the method 700 as illustrated in the flowchart of FIG. 7, comprising the following steps:

[00124] Step S701: Obtaining a modification instruction based on a manual review result, where the modification instruction includes a subtraction instruction and an addition instruction.

[00125] Step S702: Modifying the first quality inspection score using the subtraction instruction and the addition instruction to obtain the second quality inspection score.

[00126] In some embodiments, the review instruction may include displaying visual information sent to relevant reviewers immediately after the first quality inspection score is obtained. The scores and related data may be displayed in a visual format on the reviewer's computer interface. The system may aggregate the initial quality inspection scores, voice text information, and any relevant metadata into a comprehensive visual representation.

[00127] In some embodiments, the visual information may include one or more of the following:

[00128] Initial quality inspection scores: The scores assigned to each voice sample by the initial quality inspection model.

[00129] Voice text information: The transcriptions of the voice samples that were analyzed to generate the scores.

[00130] Metadata: Additional information related to each voice sample, such as the time of the call, the caller's ID, the agent's ID, etc.

[00131] Hit rules: Indications of which quality inspection rules were hit (triggered) for each voice sample.

[00132] Score adjustments: Highlighted areas or fields where reviewers can input modifications to the initial scores, such as deletions or additions.

[00133] The reviewers may compare each speech content (data packet) with the matched rules, check for any missing matched rules, and determine whether to add custom matched rules. The review result may refer to the outcome obtained through the review process. After receiving the review result information, the second speech quality inspection model or module can confirm whether the matched rules for all the text information of speech to be inspected in each speech information may be modified. If no modification is needed, the first quality inspection score for the matched rule is maintained. If modification is needed, the system identifies the modification information. If a rule judged as incorrect is required to be deleted, the score corresponding to the incorrectly matched rule is deleted based on the subtraction instruction, and the incorrect information is feedback to the second speech quality inspection model to train it, thereby improving the accuracy of quality inspection. If a missing matched rule needs to be added, the corresponding matched rule may be added after the speech information where the matched rule needs to be added based on the addition instruction, and the score corresponding to the matched rule is added to the first quality inspection score. After traversing all the text information of speech to be inspected, the final score may be taken as the second quality inspection score, which may be the final score. Manual review can improve the tolerance rate of quality inspection results, and the second speech quality inspection model can be trained based on feedback information to enhance its intelligence and accuracy in subsequent speech quality inspections.

[00134] In some embodiments, the training of the second speech quality inspection model may include one or more of the following:

[00135] Collection of review feedback which may include:

- Triggering of the first review instruction: After the first quality inspection score is generated, the system may be configured to automatically trigger the first review instruction, sending visual information to the reviewer.
- Review process: The reviewer compares the speech content with the hit rules to determine if modifications to the quality inspection score are needed. If

modifications are required, the reviewer performs delete or add operations and may provide modification reasons or suggestions.

- Feedback collection: The system collects modification operations, modification reasons, suggestions, and other feedback information from the reviewer.

[00136] Error analysis and identification which may include:

- Error identification: The system analyzes the collected feedback information to identify which quality inspection rules have been incorrectly hit or missed.
- Cause analysis: Further analysis may be conducted to determine the causes of the errors, which may include unreasonable settings of quality inspection rules, ASR speech recognition errors, or improper text processing.

[00137] Model adjustment and optimization, which may include:

- Rule adjustment: Based on the error analysis results, adjustments are made to the speech quality inspection rules. For example, if a rule is frequently hit incorrectly, the regular expression or scoring mechanism of the rule may need to be modified.
- Model retraining: The adjusted quality inspection rules are reapplied to the second speech quality inspection model, and the model is retrained using a new dataset (which may include historical data and new data). The purpose of retraining is to better adapt the model to the new quality inspection rules and data distribution.
- Parameter tuning: During the retraining process, it may also be necessary to tune the model's parameters to improve its accuracy and generalization ability.

[00138] Effectiveness evaluation and iteration which may include:

- Effectiveness evaluation: A test dataset is used to evaluate the retrained model, checking if its accuracy and generalization ability have been improved.
- Iterative Optimization: If the evaluation results show that the model's performance still does not meet expectations, the steps of rule adjustment, model retraining, and parameter tuning may need to be repeated.

[00139] In some embodiments, the processing unit 104 may be configured to generate a quality inspection report based on the second quality inspection score.

[00140] The quality inspection report may include visual report information generated based on the content information of the second quality inspection score. The content of the quality inspection report includes but is not limited to: the ability to retrieve recordings containing

specific keywords through the first quality inspection report; the capability to perform cluster analysis on high-frequency words to generate a hot word report; the ability to conduct multi-dimensional statistics on quality inspection issues to generate a quality inspection tag report; the capability to count the number of problem recordings based on the agent dimension to generate an agent's quality inspection report and other report information. Intelligent quality inspection can standardize quality inspection criteria and improve quality inspection efficiency; it seamlessly covers every customer service call, achieving 100% full quality inspection and effectively avoiding the bias of sampling methods. Relying on the first quality inspection report, it is possible to summarize the main issues of all customer service personnel and the main issues of individual customer service personnel, and address and correct them, thereby improving customer service quality.

[00141] FIG. 8 shows a generalized flow chart of a method 800 according to the present disclosure, the method comprises:

[00142] Step S801: constructing a first speech quality inspection model based on a set of speech quality inspection rules defining one or more criteria used to assess the quality of speech and a set of scoring rules defining evaluation of speech interactions based on the one or more criteria;

[00143] Step S802: configuring the first speech quality inspection model based on a task list to obtain a second speech quality inspection model, the task list comprising a set of operations or steps to be performed during a speech quality inspection;

[00144] Step S803: obtaining a data packet, the data packet comprising speech or voice converted by an automatic speech recognition (ASR) system/module into text data, and inputting the data packet into the second speech quality inspection model; and

[00145] Step S804: obtaining a first quality inspection score based on an output of the second speech quality inspection model.

[00146] The steps of the various disclosed methods are not restricted to the sequence set forth in the claims unless explicitly required. Steps may be performed in a different order, concurrently, or omitted entirely, depending on the embodiment.

[00147] It may be appreciable that the server apparatus may be used in various applications, potentially in any applications involving ASR systems, and the evaluation of output generated by ASR systems.

[00148] In some embodiments, the first quality inspection score may be obtained based on a machine learning based scoring method. The machine learning method may not fully rely on

manually set rules and one-to-one classification labels. In some embodiments, a large amount of historical speech quality inspection data can be used as a training set, and the machine learning model may be trained using one or more machine learning algorithms, such as decision trees, random forests, support vector machines in supervised learning, or neural networks in deep learning. During the training process, the model may be configured to learn the correlation rules between text features and quality inspection scores. When new speech texts to be inspected are input, the trained model may output scores according to the learned rules. This scoring method focuses more on mining potential quality assessment rules from the data, rather than explicitly performing manual rule matching and one-to-one correspondence of classification labels as in the example of steps S601 to S604.

[00149] In some embodiments, the first quality inspection score may be obtained based on a context based dynamic scoring method. Considering that the text information of speech quality inspection may comprise rich context, more complex natural language processing technologies such as semantic understanding and sentiment analysis can be adopted. By deeply analyzing the text, scores may be dynamically given to different parts of the text content. For example, in a customer service conversation, according to the context, it may be determined whether the customer service's answer to the customer's question is complete and whether there are comforting words, etc. Then, the first quality inspection score is comprehensively obtained based on these dynamic analysis results. The context based dynamic scoring method may not be limited to simple rule matching and classification label correspondence, and may instead pay more attention to the actual performance and semantic value of the text content in a specific scenario.

[00150] The methods described herein may be performed and the various processing or computation units and the devices and computing entities described herein may be implemented by one or more circuits. In an embodiment, a "circuit" may be understood as any kind of a logic implementing entity, which may be hardware, software, firmware, or any combination thereof. Thus, in an embodiment, a "circuit" may be a hard-wired logic circuit or a programmable logic circuit such as a programmable processor, e.g. a microprocessor. A "circuit" may also be software being implemented or executed by a processor, e.g. any kind of computer program, e.g. a computer program using a virtual machine code. Any other kind of implementation of the respective functions which are described herein may also be understood as a "circuit" in accordance with an alternative embodiment.

[00151] While the disclosure has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the disclosure as defined by the appended claims. The scope of the disclosure is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A server apparatus, the server apparatus comprising a processor, the processor configured to:

construct a first speech quality inspection model based on a set of speech quality inspection rules defining one or more criteria used to assess quality of speech and a set of quality inspection scoring rules defining evaluation of speech interactions based on the one or more criteria;

configure the first speech quality inspection model based on a task list to obtain a second speech quality inspection model, the task list comprising a set of operations or steps to be performed during a speech quality inspection;

obtain a data packet, the data packet comprising speech or voice converted by an automatic speech recognition (ASR) module into text data, into the second speech quality inspection model; and

obtain a first quality inspection score based on an output of the second speech quality inspection mode.

2. The server apparatus of claim 1, wherein the processor is configured to obtain a second quality inspection score based on a review of the first quality inspection score.

3. The server apparatus of claim 2, wherein the processor is configured to generate a quality inspection report based on the second quality inspection score.

4. The sever apparatus of any one of the preceding claims, wherein each task in the task list comprises a plurality of tags, each tag associated with one or more of the following: a time node information for each task to be completed, a data volume information for each task to be completed, a quality inspection classification information for each task to be completed.

5. The server apparatus of claim 4, wherein the processor is configured to generate a task allocation information based on the task list, and wherein the task allocation information comprises quality inspection time information and the quality inspection classification information.

6. The server apparatus of claim 5, wherein the processor is configured to obtain a first configuration information based on the quality inspection time information, and/or obtain a second configuration information based on the quality inspection classification information.
7. The server apparatus of claim 6, wherein the second speech quality inspection model is generated based on the first configuration information and the second configuration information.
8. The server apparatus of any one of the preceding claims, wherein the processor is configured to obtain a modification instruction based on the review of the first quality inspection score, and wherein the modification instruction includes a subtraction instruction and/or an addition instruction, and wherein the processor is further configured to modify the first quality inspection score using the subtraction instruction and/or the addition instruction to obtain the second quality inspection score.
9. The server apparatus of any one of the preceding claims, wherein the set of speech quality inspection rules comprises one or more manual speech quality inspection rules, and wherein the processor is configured to:
 - convert the one or more manual speech quality inspection rules into a text matching rule set, where the text matching rule set includes word-based rules and phrase-based rules; and
 - generate the speech quality inspection rules using the word-based rules and the phrase-based rules.
10. The server apparatus of any one of the preceding claims, wherein the processor is further configured to train the first or second speech quality inspection model.
11. The server apparatus of claim 10, wherein the processor is configured to obtain a training dataset, and wherein the training dataset comprises a historical dataset processed by an independent speech quality inspection module, the historical dataset comprising a recording text dataset and a customer feedback information dataset.

12. The server apparatus of claim 11, wherein the processor is configured to obtain a data volume value based on the recording text dataset and the customer feedback information dataset, and obtain a preset data volume value, and determine whether the data volume value is less than the preset data volume value.
13. The server apparatus of claim 12, wherein in a positive determination that the data volume value is less than the preset data volume value, the processor is configured to: obtain one or more manual quality inspection scoring rules.
14. The server apparatus of claim 9 and 12, wherein in a negative determination that the data volume value is less than the preset data volume value, the processor is configured to: obtain an automatic generation instruction for the speech quality inspection rules; preprocess the recording text dataset based on the automatic generation instruction for the set of speech quality inspection rules to obtain a recording text information, wherein the second recording text dataset includes customer recording text and customer service recording text.
15. The server apparatus of claim 14, wherein the processor is configured to encode the customer recording text and the customer service recording text using an encoding instruction to obtain an encoding result.
16. The server apparatus of claim 15, wherein the processor is further configured to generate a classification label based on a business type of the second recording text dataset, and identify the encoding result based on the classification label to generate the quality inspection scoring rules.
17. A system for speech quality inspection analysis, the system comprises
an automatic speech recognition module;
the server apparatus of any one of the preceding claims, the server apparatus arranged in data communication with the automatic speech recognition module to receive a data packet, the data packet comprising speech or voice data converted by the automatic speech recognition module into text data; and
wherein the server apparatus is configured to display visual information, the visual information comprising the first quality inspection score.

18. The system of claim 17, wherein the visual information further comprises voice text information, and metadata.
19. A method for determining a parameter of speech data packets, the method comprises:
 - constructing a first speech quality inspection model based on a set of speech quality inspection rules defining one or more criteria used to assess the quality of speech and a set of scoring rules defining evaluation of speech interactions based on the one or more criteria;
 - configuring the first speech quality inspection model based on a task list to obtain a second speech quality inspection model, the task list comprising a set of operations or steps to be performed during a speech quality inspection;
 - obtaining a data packet, the data packet comprising speech or voice converted by an automatic speech recognition (ASR) module into text data, inputting the data packet into the second speech quality inspection model; and
 - obtaining a first quality inspection score based on an output of the second speech quality inspection model.
20. A non-transitory computer-readable medium storing computer executable code comprising instructions for processing data according to the method of claim 19.