TITLE OF INVENTION: AUDIO TRANSCRIPTION METHOD AND DEVICE FOR ELIMINATING HALLUCINATION EFFECT

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to and the benefit of Korean Patent Application No. 10-20xx-0xxxxxx filed in the Korean Intellectual Property Office on MONTH xx, 20xx, the entire contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

(a) Field of the Invention

[0002] The present invention relates to an audio transcription method and device for eliminating a hallucination effect, and more specifically, to an invention capable of eliminating the hallucination effect that occurs during audio transcription by preprocessing input audio data to remove silent sections.

(b) Description of the Related Art

[0003] Audio transcription refers to a process of converting a voice recording or an audio file into a text format. In this process, voice content included in an audio file can be converted into text in an audible language and recorded.

[0004] Such audio transcription is based on voice recognition technology and can be mainly performed through voice recognition software or services.

[0005] To achieve this, various technologies and algorithms can be used to process audio files or recorded voice data and convert the speech content included therein into text.

[0006] However, in such audio transcription technology, there is a problem that a hallucination effect occurs, and this hallucination effect can be defined as a case where an audio transcription system generates an inaccurate result for a given voice input.

[0007] In particular, the hallucination effect is implemented as an error such as repetition of data with the same content in text data, which is the output data of the audio transcription system, and this can be generated due to the input audio data.

[0008] For example, if noise or background noise is mixed in the input audio data, there is a problem that the transcription system cannot properly process it and recognizes the noise as text and outputs an incorrect result.

SUMMARY OF THE INVENTION

[0009] Therefore, an audio transcription method and device for eliminating a hallucination effect according to one embodiment of the disclosed invention is created to solve the problems of the above-described conventional technology, and more specifically, an audio transcription method and device capable of preventing transcription errors by removing a silent section from input audio data may be provided.

[0010] In addition, the audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention can provide an audio transcription method and device that performs audio transcription in various ways until the hallucination effect is completely eliminated by determining whether text data, which is the final output data, includes the hallucination effect.

[0011] In addition, the audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention can provide an audio transcription method and device that can more effectively remove a silent section from input data by using various types of silence determination modules.

[0012] An audio transcription device for eliminating a hallucination effect, comprise a voice acquisition unit that receives audio data from a user terminal and a processor that receives audio data from the voice acquisition unit, wherein the processor comprises a silence detection module that detects a silent section included in the audio data and removes the silent section from the audio data to leave only a voice section, a text output module that divides the audio data, from which the silent section has been removed, into audio block data of first units and transcribes the divided audio block data to output text data and a hallucination determination module that determines a repeated error included in the text data.

[0013] The audio transcription device for eliminating the hallucination effect further comprise a noise removal module that removes noise from the audio data

and transmits the audio data to the silence detection module when the hallucination determination module determines that the text data includes a preset number of repeated errors or more.

[0014] The audio transcription device for eliminating the hallucination effect according to claim 1, wherein

[0015] The hallucination determination module determines that the text data includes a preset number of repeated errors or more, the text output module divides the audio data, from which the silent section has been removed, into audio block data of second units different from the first units and transcribes the divided audio block data to output text data.

[0016] The silence detection module comprises a first silence detection module that divides the audio data into a plurality of pieces of chunking data of first units and determines whether each of the plurality of pieces of chunking data is a silent section or a voice section and a second silence detection module that divides the audio data into a plurality of pieces of chunking data of second units and determines whether each of the plurality of pieces of chunking data is a silent section or a voice section.

[0017] Th first silence detection module removes chunking data corresponding to a silent section among the plurality of pieces of chunking data and merges chunking data corresponding to a remaining voice section.

[0018] An audio transcription method for eliminating a hallucination effect, comprise receiving audio data, detecting a silent section included in the audio

data, removing the silent section from the audio data to leave only a voice section, dividing the audio data, from which the silent section has been removed, into audio block data of first units and transcribing audio based on the divided audio block data and outputting text data.

[0019] The audio transcription method for eliminating the hallucination effect further comprise determining a repeated error included in the text data and removing noise from the audio data before detecting the silent section when it is determined that the text data includes a preset number of repeated errors or more.

[0020] The audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention has an advantage of preventing transcription errors by removing a silent section from input audio data.

[0021] In addition, the audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention has an advantage of performing audio transcription in various ways until the hallucination effect is completely eliminated by determining whether text data, which is the final output data, includes the hallucination effect.

[0022] In addition, the audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention has an advantage of being able to more effectively remove a silent section from input data by using various types of silence determination modules.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] FIG. 1 is a diagram illustrating an audio transcription system for eliminating a hallucination effect according to one embodiment of the disclosed invention.

[0024] FIG. 2 is a diagram illustrating the configuration of an audio transcription device for eliminating the hallucination effect according to one embodiment of the disclosed invention.

[0025] FIG. 3 is a diagram illustrating a detailed configuration of the processor 130 illustrated in FIG. 2.

[0026] FIG. 4 is a diagram illustrating an artificial neural network module utilized to output text data from audio data in the audio transcription device for eliminating the hallucination effect according to one embodiment of the disclosed invention.

[0027] FIG. 5 is a diagram showing audio data and a spectrogram of audio to which an audio transcription model according to the disclosed invention is not applied.

[0028] FIG. 6 is a diagram showing a spectrogram of audio data and audio to which an audio transcription model according to one embodiment of the disclosed invention is applied.

[0029] FIG. 7 is a diagram illustrating the overall concept of an audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention.

[0030] FIG. 8 is a flowchart illustrating an audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0031] The embodiments described in this specification and the configurations shown in the drawings are only preferred examples of the disclosed invention, and at the time of filing this application, there may be various modifications that can replace the embodiments and drawings in this specification.

[0032] In addition, the same reference numbers or symbols shown in each drawing of this specification indicate parts or components that perform substantially the same function.

[0033] Additionally, the terms used herein are used to describe embodiments and are not intended to limit and/or limit the disclosed invention. Singular expressions include plural expressions unless the context clearly dictates otherwise.

[0034] In this specification, terms such as “comprise” or “have” is intended to designate the presence of features, numbers, steps, operations, components, parts, or combinations thereof described in the specification, but are not intended to indicate the presence of one or more other features. The existence or addition of numbers, steps, operations, components, parts, or combinations thereof is not excluded in advance.

[0035] In addition, terms including ordinal numbers such as “first”, “second”, etc. used in this specification may be used to describe various components, but the components are not limited by the terms, and the terms are used only for the purpose of distinguishing one component from another.

[0036] For example, a first component may be named a second component, and similarly, the second component may also be named a first component without departing from the scope of the present invention. The term “and/or” includes any combination of a plurality of related stated items or any of a plurality of related stated items.

[0037] Hereinafter, embodiments according to the present invention will be described in detail with reference to the attached drawings.

[0038] FIG. 1 is a diagram illustrating an audio transcription system for eliminating a hallucination effect according to one embodiment of the disclosed invention.

[0039] Referring to FIG. 1, the audio transcription system for eliminating the hallucination effect according to one embodiment of the disclosed invention may include an audio transcription device 100 and a user terminal 200.

[0040] For example, the audio transcription device 100 according to the present invention may be implemented as various devices, for example, it may be implemented as a user terminal, which is a client terminal, or it may be implemented as a server that acquires information from a client and provides such services.

[0041] In the present invention, the server means a typical server, and the server is a computer hardware on which a program is executed, and may monitor or control the entire network, such as printer control or file management, or support connection with other networks through a mainframe or public network and sharing of software resources such as data, programs, and files, or hardware resources such as modems, faxes, printer sharing, and other equipment.

[0042] In addition, the user terminal 200 according to one embodiment of the disclosed invention may correspond to a computing device that may transmit and receive data using a network and execute various kinds of software.

[0043] Here, a network means a connection structure that enables information exchange between each node, such as a plurality of terminals and servers, and examples of such networks include a local area network (LAN), a wide area network (WAN), the Internet (WWW: World Wide Web), a wired and wireless data communication network, a telephone network, and a wired and wireless television communication network. Examples of wireless data communication networks include, but are not limited to, 3G, 4G, 5G, 3GPP (3rd Generation Partnership Project), 5GPP (5th Generation Partnership Project), LTE (Long Term Evolution), WIMAX (World Interoperability for Microwave Access), Wi-Fi, the Internet, LAN (Local Area Network), Wireless LAN (Wireless Local Area Network), WAN (Wide Area Network), PAN (Personal Area Network), RF (Radio Frequency), Bluetooth (Bluetooth) network, NFC (Near-Field Communication) network, satellite broadcasting network, analog broadcasting network, and DMB (Digital

Multimedia Broadcasting) network.

[0044] The user terminal 200 may store an operating system program for booting and operation.

[0045] For example, the operating system program may include Android OS, iOS (iPhone OS), Tizen OS, BlackBerry OS, Windows Phone OS, and the like.

[0046] In addition, the user terminal 200 may be implemented as a plurality of terminal devices including a processor 130 so that such an algorithm may be realized, and may be implemented as a computer, a smart pad, a laptop, or a portable terminal, for example.

[0047] In addition, when the audio transcription device 100 according to the disclosed invention is implemented in the user terminal 200, the user terminal 200 may perform transcription by removing a silent section after receiving audio data from a user, dividing the input data of the text output into an appropriate length, and outputting text data from the audio data.

[0048] On the other hand, when the audio transcription device 100 according to the present invention is implemented as a server, the user terminal 200 may transmit audio data input by the user to the server, and the server may process and divide the audio data based on the audio data to output text data.

[0049] FIG. 2 is a diagram illustrating the configuration of an audio transcription device for eliminating the hallucination effect according to one embodiment of the disclosed invention. FIG. 3 is a diagram illustrating a detailed configuration of the processor 130 illustrated in FIG. 2. FIG. 4 is a diagram illustrating an artificial

neural network module utilized to output text data from audio data in the audio transcription device for eliminating the hallucination effect according to one embodiment of the disclosed invention.

[0050] Referring to FIG. 2, the audio transcription device for eliminating the hallucination effect according to one embodiment of the disclosed invention may include an input unit 110 for receiving various pieces of data from a user, a voice acquisition unit 120 configured to acquire audio data input from the user terminal 200, a processor 130 for preprocessing and combining various pieces of information acquired from the input unit 110 and the voice acquisition unit 120 and converting them into text, a memory 140 for storing the user's audio data and the like, and an interface unit 150 for displaying various pieces of information including text data.

[0051] The input unit 110 may receive various pieces of data from the user. For example, the data input from the user may include information such as the user's personal information and age.

[0052] The user-related data thus input may be stored in the memory 140 or transmitted to a server through a communication module 136 of the processor 130.

[0053] The voice acquisition unit 120 may receive the user's audio data from the user terminal 200, acquire the audio data, and manage the audio data.

[0054] Specifically, the audio data may be information obtained by the user directly inputting voice through an input means such as a microphone (not shown).

[0055] The audio data acquired in this way may be converted into a digital voice file by the processor 130 described below, and at the same time, an identifier may be assigned to each user.

[0056] In addition, the voice file acquired by the voice acquisition unit 120 may be stored in the memory 140.

[0057] The processor 130 may preprocess the user's audio data into information that may be utilized in an artificial neural network module used for STT, and convert the user's audio data into text data and output the text data.

[0058] Specifically, the processor 130 may convert input audio data into text data through STT technology, and may perform a process of removing a silent section from the audio data a plurality of times so that the output text data does not include a hallucination effect, or may perform a process of removing noise from the audio data. Details related to this will be described later.

[0059] Referring to FIG. 3, the processor 130 may include a noise removal module 131, a first silence detection module 132, a second silence detection module 133, a text output module 134, a hallucination determination module 135, and a communication module 136.

[0060] More specifically, the noise removal module 131 of the audio transcription device 100 according to one embodiment of the disclosed invention may remove noise from the audio data input from a user and transmit the audio data to the silence detection module.

[0061] For example, the noise removal module 131 of the present invention

may be configured to process audio in the waveform domain.

[0062] Specifically, the noise removal module 131 of the present invention may be implemented using a noise removal model based on a casual demux, instead of using a predetermined threshold for power or volume level.

[0063] More specifically, according to the noise removal module 131 of the present invention, when audio data X is given, audio data may be configured as clean audio data Y damaged by a background signal N such that the equation X = Y+N is satisfied.

[0064] At this time, the noise removal module 131 of the present invention may be configured to use the audio data X as input and reconstruct clean audio Y as output in order to reduce loss in both the time domain and the frequency domain between the reconstructed signals Y' and Y.

[0065] After that, the noise removal module 131 of the present invention may be implemented to automatically fine-tune noise from input audio data instead of a threshold manually set by a user.

[0066] In addition, as described below, the noise removal module 131 of the disclosed invention may be activated to more effectively remove silent sections from audio data when the hallucination determination module 135 determines that there is a repeated error in the final output text data.

[0067] Specifically, the noise removal module 131 of the disclosed invention may remove noise from audio data and transmit the audio data to the silence detection module when the hallucination determination module 135 determines

that the text data includes ten or more repeated errors. Details related to this will be described later.

[0068] In addition, the silence detection module of the audio transcription device 100 according to one embodiment of the disclosed invention may include a first silence detection module 132 and a second silence detection module 133.

[0069] More specifically, the silence detection module of the disclosed invention may be configured to detect a silent section included in audio data and remove the silent section from the audio data to leave only a voice section.

[0070] However, the silence detection module of the disclosed invention may include a first silence detection module 132 used as the default and a second silence detection module 133 additionally activated by the hallucination determination module 135.

[0071] For example, the first silence detection module 132 may be configured to divide audio data into a plurality of pieces of chunking data of 5 seconds as the first units and determine whether each of the plurality of pieces of chunking data is a silent section or a voice section.

[0072] Specifically, the first silence detection module 132 may divide the initial audio data into a plurality of pieces of chunking data of 5 seconds based on a stride of 2.5 seconds.

[0073] After that, the first silence detection module 132 may output the probability that each piece of chunking data of 16 ms is a voice section or a silent section.

[0074] In addition, the first silence detection module 132 may determine whether each piece of chunking data is a silent section or a voice section using a threshold of 0.5 based on the output results of each piece of chunking data of 16 ms.

[0075] For example, in one chunking data, if a continuous section of 100 ms or more is determined to be a “voice section”, the section may be displayed as a “voice section”.

[0076] In addition, the display of “voice section” may be maintained until a continuous section of 100 ms or more is determined to be a “silent section”.

[0077] After that, among the “voice section” and the “silent section” in the chunking data, a section having a value of 0.5 or more, which is a threshold, determines the type of chunking data.

[0078] For example, if the “voice section” in one chunking data is 0.6 and the “silent section” is 0.4, the chunking data may be classified as a “voice section”.

[0079] After that, the first silence detection module 132 of the disclosed invention may remove the chunking data corresponding to the “silent section” from among the plurality of pieces of chunking data, merge the chunking data corresponding to the remaining “voice section”, and reconstruct it as input data of the text output module 134.

[0080] In addition, the second silence detection module 133 may be configured to divide the audio data into a plurality of pieces of chunking data of 30 ms as the second units, and determine whether each of the plurality of pieces of chunking

data is a silent section or a voice section.

[0081] Specifically, the first silence detection module 133 may divide the initial audio data into a plurality of pieces of chunking data of 30 ms based on a stride of 9.6 ms.

[0082] After that, the second silence detection module 133 may determine whether each piece of chunking data is a voice section or a silent section. This is a method of detecting a silent section using an audio wave different from the first silence detection module 132.

[0083] For example, in a plurality of pieces of chunking data, if a continuous section of 250 ms or longer is determined to be a “voice section”, the corresponding plurality of chunking data sections may be displayed as a “voice section”.

[0084] In addition, the display of this “voice section” may be maintained until a continuous section of 100 ms or longer is determined to be a “silent section”.

[0085] After that, the second silence detection module 133 of the disclosed invention may remove chunking data corresponding to a “silent section” from among the plurality of pieces of chunking data, merge the chunking data corresponding to the remaining “voice section”, and reconstruct it as input data of the text output module 134.

[0086] In addition, the text output module 134 of the audio transcription device 100 according to one embodiment of the disclosed invention may divide audio data edited by the silence detection module into a plurality of pieces of audio

block data, and transcribe each of the pieces of divided audio block data to output text data.

[0087] For example, the text output module 134 of the audio transcription device 100 according to one embodiment of the disclosed invention may receive audio data edited by the silence detection module as input data and, in some cases, divide it into audio block data of 30 seconds as the first units or audio block data of 5 seconds as the second units.

[0088] Specifically, the text output module 134 of the audio transcription device 100 according to one embodiment of the disclosed invention may, by default, receive audio data edited by the silence detection module as input data and divide audio data, from which silent sections are removed into audio block data of 30 seconds.

[0089] However, the text output module 134 of the audio transcription device 100 according to one embodiment of the disclosed invention may divide the audio data with the silent section removed into audio block data of 5 seconds instead of 30 seconds when the hallucination determination module, which will be described later, determines that the text data includes ten or more repeated errors.

[0090] Through this, the text output module 134 of the disclosed invention has a technical effect of reducing the possibility of hallucinations occurring due to noise being introduced.

[0091] In addition, the hallucination determination module 135 of the audio transcription device 100 according to one embodiment of the disclosed invention

may be configured to determine repeated errors included in the text data output by the text output module 134.

[0092] For example, if it is determined that the number of repeated errors included in the text data finally output by the hallucination determination module 135 of the disclosed invention is less than a preset number of 10, the audio transcription device 100 of the disclosed invention may terminate the audio transcription process without additional steps.

[0093] However, if it is determined that the number of repeated errors included in the text data finally output by the hallucination determination module 135 of the disclosed invention is more than the preset number of 10, the audio transcription device 100 of the disclosed invention may eliminate the hallucination effect by processing the audio data input to the first silence detection module 132 using the noise detection module in addition to the audio transcription process performed by the first silence detection module 132 and the text output module 134.

[0094] Alternatively, if it is determined that the number of repeated errors included in the text data finally output by the hallucination determination module 135 of the disclosed invention is more than the preset number of 10, the audio transcription device 100 of the disclosed invention may perform a more detailed silent section determination in addition to the silent section determined by the first silence detection module 132 using the second silence detection module 133.

[0095] Alternatively, if it is determined that the number of repeated errors

included in the text data finally output by the hallucination determination module 135 of the disclosed invention is more than the preset number of 10, the audio transcription device 100 of the disclosed invention may process the data input to the first silence detection module 132 using both the noise detection module and the second silence detection module 133, and perform a silent section determination in addition to the silent section determined by the first silence detection module 132.

[0096] Alternatively, if it is determined that the repeated error included in the text data finally output by the hallucination determination module 135 of the disclosed invention is more than the preset number of 10, the audio transcription device 100 of the disclosed invention may divide the audio block data used in the final audio transcription into 5-second units instead of the first units of 30 seconds, thereby more completely eliminating the hallucination effect.

[0097] In addition, the communication module 136 of the audio transcription device 100 according to one embodiment of the disclosed invention may be configured to transmit and receive information between the user terminal 200 and the audio transcription device 100.

[0098] For example, such a communication module 136 may include at least one of a short-range communication module, a wired communication module, and a wireless communication module.

[0099] Referring to FIG. 4, the artificial neural network module utilized in the audio transcription device 100 of the disclosed invention may use preprocessed

audio data as input information and text data converted into text from the audio data as output information.

[0100] In particular, various types of known artificial neural networks may be used in the artificial neural network module according to the present invention, and a convolutional neural network (CNN) may be used as a representative example.

[0101] However, the artificial neural network module according to the present invention is a model for establishing a classification algorithm, and various artificial intelligence models may be utilized in addition to the CNN described above.

[0102] For example, the artificial neural network module according to the present invention may utilize a transformer model for establishing a classification algorithm. Transformer is a deep learning model mainly used in natural language processing (NLP), and is a model proposed in the paper “Attention is All You Need” published in 2017.

[0103] The transformer model has strengths in processing sequence data, and unlike conventional recurrent neural networks (RNNs), it may model interactions between words in a sentence using a self-attention mechanism.

[0104] In addition, the artificial neural network module according to the present invention may utilize the bidirectional encoder representations from transformers (BERT) model to establish a classification algorithm.

[0105] The BERT model is a pre-training model for natural language processing

announced by Google in 2018, and is a model based on the transformer model.

[0106] In addition, the artificial neural network module according to the present invention may utilize a support vector machine (SVM) model to establish a classification algorithm.

[0107] SVM is a supervised learning algorithm which is mainly used for classification tasks and may be used to find boundaries between classes by mapping data to a high-dimensional space.

[0108] As will be described later, the artificial neural network module according to the present invention may perform learning by comparing output information and reference data when performing learning, but may perform learning using different types of input information.

[0109] Therefore, the artificial neural network module according to the disclosed invention may include a learning session that outputs output information based on input information, and an inference session that infers output information based on input information.

[0110] The artificial neural network module according to the disclosed invention may perform learning based on output information and reference information when performing learning.

[0111] Here, the reference information may be reference data generated artificially.

[0112] CNN is a deep learning model specialized in detecting local patterns from input data and extracting complex features, and may perform vision tasks

such as image classification, object detection, and segmentation.

[0113] CNN may include a convolution layer that extracts local features using a filter for input data, a pooling layer that reduces the spatial size of the feature map extracted from the convolution layer to reduce the amount of calculation and maintain abstracted features, an activation function used to calculate the output of each neuron, and a fully connected layer located at the end of the CNN to perform the final classification or prediction based on the extracted features.

[0114] The fully connected layer receives the output of the previous layer as input and outputs a score for each class and may generally obtain a probability distribution for each class using a softmax function.

[0115] Therefore, the artificial neural network module according to the disclosed invention is excellent in correcting errors by combining the results of multiple analysis models, so that the performance may be further improved.

[0116] In addition, the artificial neural network module according to the disclosed invention has excellent generalization and scalability, so it has the strength of being able to perform information processing for various data types and sizes flexibly.

[0117] FIG. 5 is a diagram showing audio data and a spectrogram of audio to which an audio transcription model according to the disclosed invention is not applied. FIG. 6 is a diagram showing a spectrogram of audio data and audio to which an audio transcription model according to one embodiment of the disclosed invention is applied.

[0118] Referring to FIGS. 5 and 6, it may be confirmed that the audio data and the spectrogram of audio to which an audio transcription model according to the disclosed invention is not applied have more noise than the audio data and the spectrogram of audio to which the audio transcription model is applied.

[0119] Specifically, as shown in FIG. 6, it may be confirmed that the audio processed by the noise removal module 131 of the audio transcription device according to the disclosed invention includes a detailed signal with reduced noise.

[0120] This may also be confirmed in the spectrogram of FIG. 6, and specifically, it may be confirmed that the noise power in the high-frequency region is reduced.

[0121] FIG. 7 is a diagram illustrating the overall concept of an audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention. FIG. 8 is a flowchart illustrating an audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention.

[0122] Referring to FIG. 7, an audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention may basically receive audio data, determine a silent section in the input audio data, and merge audio data, from which the silent section has been removed.

[0123] After that, the disclosed invention may perform audio transcription by dividing the merged audio data, from which the silent section has been removed, into audio block data of 30 seconds and extracting text data from the audio data

based on the audio block data.

[0124] After that, the disclosed invention may determine whether the extracted text data includes hallucination.

[0125] Specifically, if the text data includes ten or more repeated errors, the disclosed invention may perform an audio preprocessing process, including noise removal or silent section removal, again.

[0126] Alternatively, if the text data includes less than ten repeated errors, the disclosed invention may determine that there is almost no hallucination and terminate the audio transcription process.

[0127] In addition, if the text data includes ten or more repeated errors and the audio preprocessing process, including noise removal or silent section removal, is performed again, the disclosed invention may preprocess the audio differently than before.

[0128] For example, when the text data includes ten or more repeated errors and the audio preprocessing process, including noise removal or silent section removal, is performed again, the disclosed invention may perform a first step to remove noise from audio data to be input to the first silence detection module 132 by using the noise removal module 131.

[0129] In addition, when the text data includes ten or more repeated errors and the audio preprocessing process, including noise removal or silent section removal, is performed again, the disclosed invention may perform a second step to remove noise from audio data to be input to the first silence detection module

132 by using the second silence detection module 133 to additionally determine a silent section in addition to the first silence detection module 132.

[0130] In addition, when the text data includes ten or more repeated errors and the audio preprocessing process, including noise removal or silent section removal, is performed again, the disclosed invention may perform more effective audio data preprocessing as a third step by using both the noise removal module 131 and the second silence detection module 133.

[0131] In addition, when the text data includes more than ten repeated errors and the audio preprocessing process, including noise removal or silent section removal, is performed again, the disclosed invention may perform a final fourth step to divide the audio data, from which the silent section has been completely removed into audio block data of 5 seconds as the second units, rather than the existing 30-second units as the first units, thereby minimizing the influence of unnecessary noise that may be included during text conversion.

[0132] Referring to FIG. 8, an audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention may include an audio data receiving step (S110) of receiving audio data.

[0133] More specifically, the audio data receiving step may include a step in which a user inputs audio data into the audio transcription device 100 through the user terminal 200.

[0134] After that, the audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention may

include a silence detection step (S120) of detecting a silent section included in the received audio data.

[0135] In addition, the audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention may include a silent section removal step (S130) of removing the detected silent section from the audio data.

[0136] The silence detection step and the silent section removal step may be performed by the silence detection module of the disclosed invention.

[0137] A detailed description of the silence detection module will be omitted because it has already been described.

[0138] In addition, the audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention may include an audio dividing step (S140) of dividing audio data, from which a silent section has been removed, into audio block data of preset arbitrary units.

[0139] More specifically, the audio dividing step (S140) of the disclosed invention may include a step of dividing audio data into 30-second sections using the text output module 134.

[0140] In addition, the audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention may include a text output step (S150) of transcribing audio based on the divided audio block data and outputting text data.

[0141] In addition, the audio transcription method for eliminating the

hallucination effect according to one embodiment of the disclosed invention may include a hallucination determination step (S160) of determining the number of repeated errors included in the output text data.

[0142] After that, the audio transcription method for eliminating the hallucination effect according to one embodiment of the disclosed invention may include a repetition step (S170) of repeatedly performing audio preprocessing through an additional process if the number of repeated errors determined in the hallucination determination step is greater than or equal to a threshold.

[0143] Specifically, the repetition step (S170) may include a step of repeatedly performing audio preprocessing through an additional process if it is determined in the hallucination determination step that the text data includes ten or more repeated errors.

[0144] For example, the repetition step (S170) may include a noise removal step of removing noise from audio data using the noise removal module 131 before the silence detection step.

[0145] In addition, the repetition step (S170) may include a step of additionally determining a silent section in the audio data by using the second silence detection module 133 together with the first silence detection module 132.

[0146] In addition, the repetition step (S170) may include a step of removing noise from audio data using both the noise removal module 131 and the second silence detection module 133 and inputting the audio data to the first silence detection module 132, and additionally determining a silent section through the

second silence detection module 133 after the silent section determination is performed by the first silence detection module 132.

[0147] In addition, the repetition step (S170) may include a step of dividing the final audio data, from which the silent section has been removed, into audio block data of 5 seconds instead of 30 seconds and inputting the data to the text output module 134.

[0148] Through this, the disclosed invention has the technical effect of reducing the false positive rate in detecting malicious applications through complementary comparison of the market crawling module 120 and the threat detection module 110.

[0149] The audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention has an advantage of preventing transcription errors by removing a silent section from input audio data.

[0150] In addition, the audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention has an advantage of performing audio transcription in various ways until the hallucination effect is completely eliminated by determining whether text data, which is the final output data, includes the hallucination effect.

[0151] In addition, the audio transcription method and device for eliminating the hallucination effect according to one embodiment of the disclosed invention has an advantage of being able to more effectively remove a silent section from input

data by using various types of silence determination modules.

[0152] The signature-based malicious application detection method and device according to one embodiment has the technical advantage of being able to quickly and accurately detect malicious applications through various verification methods.

[0153] In addition, the signature-based malicious application detection method and device according to one embodiment has the technical advantage of being able to immediately reflect malicious applications deleted from the market in the signature database by crawling the vendor market app posting history.

[0154] In addition, the signature-based malicious application detection method and device according to one embodiment uses the developer characteristics and certificate characteristics of the application for machine learning and provides a technical technology that can effectively protect users from malicious applications through regular signature database updates. There are advantages.

[0155] In the above, specific embodiments are shown and described. However, it is not limited to the above-mentioned embodiments, and those skilled in the art can make various changes without departing from the gist of the technical idea of the invention as set forth in the claims below.

WHAT IS CLAIMED IS:

[CLAIM 1] An audio transcription device for eliminating a hallucination effect, comprising:

a voice acquisition unit that receives audio data from a user terminal; and

a processor that receives audio data from the voice acquisition unit, wherein

the processor comprises:

a silence detection module that detects a silent section included in the audio data and removes the silent section from the audio data to leave only a voice section;

a text output module that divides the audio data, from which the silent section has been removed, into audio block data of first units and transcribes the divided audio block data to output text data; and

a hallucination determination module that determines a repeated error included in the text data.

[CLAIM 2] The audio transcription device for eliminating the hallucination effect according to claim 1, further comprising:

a noise removal module that removes noise from the audio data and transmits the audio data to the silence detection module when the hallucination determination module determines that the text data includes a preset number of

repeated errors or more.

[CLAIM 3] The audio transcription device for eliminating the hallucination effect according to claim 1, wherein

when the hallucination determination module determines that the text data includes a preset number of repeated errors or more, the text output module divides the audio data, from which the silent section has been removed, into audio block data of second units different from the first units and transcribes the divided audio block data to output text data.

[CLAIM 4] The audio transcription device for eliminating the hallucination effect according to claim 1, wherein

the silence detection module comprises:

a first silence detection module that divides the audio data into a plurality of pieces of chunking data of first units and determines whether each of the plurality of pieces of chunking data is a silent section or a voice section; and

a second silence detection module that divides the audio data into a plurality of pieces of chunking data of second units and determines whether each of the plurality of pieces of chunking data is a silent section or a voice section.

[CLAIM 5] The audio transcription device for eliminating the hallucination effect according to claim 4, wherein

the first silence detection module removes chunking data corresponding

to a silent section among the plurality of pieces of chunking data and merges chunking data corresponding to a remaining voice section.

[CLAIM 6] An audio transcription method for eliminating a hallucination effect, comprising:

receiving audio data;

detecting a silent section included in the audio data;

removing the silent section from the audio data to leave only a voice section;

dividing the audio data, from which the silent section has been removed, into audio block data of first units; and

transcribing audio based on the divided audio block data and outputting text data.

[CLAIM 7] The audio transcription method for eliminating the hallucination effect according to claim 6, further comprising:

determining a repeated error included in the text data; and

removing noise from the audio data before detecting the silent section when it is determined that the text data includes a preset number of repeated errors or more.