

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 July 2002 (11.07.2002)

PCT

(10) International Publication Number
WO 02/054333 A2

- (51) International Patent Classification⁷: **G06K**
- (21) International Application Number: PCT/IL01/01221
- (22) International Filing Date:
31 December 2001 (31.12.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
140673 1 January 2001 (01.01.2001) IL
- (71) Applicant (*for all designated States except US*): **POLY INFORMATION LTD.** [IL/IL]; Menachem Begin Road 48, 66184 Tel Aviv (IL).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **ALT, Ofer** [IL/IL]; Hanevatim Street 7, 52960 Ramat Efal (IL). **RAPOPORT, Simon** [IL/IL]; Mifratz Shlomo 4/27, 58498 Holon (IL). **SHAMIR, Oren** [IL/IL]; Admonit 33/13, 71473 Lod (IL). **KNYAZHANSKY, Ilya** [IL/IL]; Rehovot Blvd., 26/5, 77640 Ashdod (IL).
- (74) Agents: **LUZZATTO, Kfir** et al.; Luzzatto & Luzzatto, P.O. Box 5352, 84152 Beer-Sheva (IL).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: A METHOD AND SYSTEM FOR IMPROVED SPEECH RECOGNITION

(57) Abstract: Method and system for improving speech recognition. A Speech Recognition (SR) system is used for outputting a sentence that matches an input sentence of a user. The SR system comprises a plurality of predetermined sentences that are associated with a specific subject, and a set of predetermined number of N sentences, selected from the plurality of predetermined sentences, having the highest level of match to the input sentence. The SR system has a predetermined threshold for the level of match, beyond which, the sentence, from the set, that has the highest level of match, is output as a recognized input sentence. A verbal input sentence is received from the user in the SR system and a weight is assigned for the level of match for each sentence from the plurality predetermined sentences according to the content of the input sentence. N sentences having the highest weight are selected from the plurality. If the weight of at least one of the N sentences is higher than the threshold, the sentence having the highest weight is output as the recognized input sentence. If the weight of each selected sentence is lower than the threshold, the weight of each selected sentence is varied according to different predetermined matching criteria. If the varied weight of at least one of the N sentences is higher than the threshold, the sentence having the highest varied weight is output as the recognized input sentence, otherwise an indication that corresponds to unrecognized input sentence is provided.

WO 02/054333 A2

A METHOD AND SYSTEM FOR IMPROVED SPEECH RECOGNITION

Field of the Invention

The present invention relates to the field of human-machine interfaces. More particularly, the present invention relates to a method and system for improving the accuracy, reliability and probability of voice recognition in human-machine interfaces.

Background of the Invention

In recent years, efforts were made to provide technologies for allowing vocal communication between human and machines. The ability to communicate in such a way has many advantages in several fields. The advantages of communicating with computerized systems by using voice as natural language input are considered as common knowledge today.

In recent years, several technological attempts have been made to access computerized systems by using human voice with different products, such as Voice eXtensible Markup Language™ (VoiceXML™) by VoiceXML™ Forum (founded by AT&T, IBM, Lucent Technologies and Motorola; the Forum web site is: <http://www.voicexml.org/>), Nuance Speech Recognition system 7.0 (Nuance Communications 2000, Menlo Park, CA, USA) etc. Such technological attempts are, among others, Voice Recognition (VR), Automatic Speech Recognition (ASR), and Text-To-Speech conversion.

In VR technology, the system tries to recognize the voice of the user and to react according to the user's orders. The problem with such technology is that the computer should "learn" variations in the tone of each user's voice in lab condition, in order to accurately identify the content of the voice expressed by the user and to correctly interpret his meaning. However, this technology is limited, since the voice recognition system become

efficient only after a "training" process, which should be repeated for each different user.

In ASR technology, the system is able to recognize the user's voice without a "training" process in lab condition. ASR technology has a set of sentences generated, for example, from databases consist Names, Addresses and Numbers that are associated with specific subject. Selected sentences from this set are compared with the vocal input from a user, and if the level of match between one of the sentences and the input sentence reaches a predetermined value, this sentence is output from the ASR system. Directed dialogue applications employ ASR technology to guide the conversation with the user, and wait for specific answers from the user. An ASR system receives sentences in human language (voice) as an input, and selects N-best sentences (selected from a plurality of sentences which are generated in advanced, using words from the databases and may represent the sentence that was probably said by the user), according to the user's input sentence (where N can be any predefined integer and positive number). Each of the N-best sentences has a corresponding "weight" (i.e., the probability percentage that this sentence has actually been said by the user), which defines the level of match with the sentence said by the user. A threshold "weight" is predefined in the ASR as well, in order to decide, which sentence (from the N-best) will be the output of the ASR. The sentence, among the N-best sentences, that has the highest weight beyond the threshold, will be output from the ASR (i.e., the most matched sentence, among the N-best, to the sentence said by the user). However, if no sentence among the N-best has passed the threshold, than the ASR will fail to provide an output. Current ASR capability to provide an accurate output sentence is limited. Therefore, the threshold level should be reduced, in order to output a sentence, but such reduction may result in inaccurate answers. Sometimes, non-of the N-best sentences passes the threshold, and therefore the ASR provides no output.

Another technology that has been developed is the Interactive Voice Reply (IVR), which is actually a menu that allows the user to choose between two or more possibilities in each stage of the conversation. This technology is also limited due to the fact that the conversation flow is restricted to the options offered by the menus and the way they are structured.

All major breakthroughs in recent years have evolved around a module called Speech-To-Text (STT) and Text-To-Speech (TTS). This core technology acts as a translator between a human voice and the written text in the computer. For example, when a person says "happy", it is translated by the module from the acoustic environment to text in a computer. However, the SST and TTS modules do not have any intelligence, but are simple translators between the acoustic environment and the written computer environment. The main breakthrough recently has been around the accuracy and reliability of such technologies. According to this technology of STT and TTS, several applications have been developed.

Critical prerequisite parameters for human conversation are the accuracy and confidence of the conversation accessories (i.e., hearing and saying). State of the art voice recognition technologies and applications lack these important conversation parameters.

All the methods described above have not yet provided satisfactory solutions to the problem of recognition voice with high reliability.

It is an object of the present invention to provide a method for improving the of voice recognition capability.

It is another object of the present invention to provide a method for increasing the accuracy of voice recognition.

Other objects and advantages of the invention will become apparent as the description proceeds.

Summary of the Invention

The present invention is directed to a method for improving speech recognition. A Speech Recognition (SR) system for outputting a sentence that matches an input sentence of a user, the SR system comprises a plurality of predetermined sentences that are associated with a specific subject, and a set of predetermined number of N sentences, selected from the plurality of predetermined sentences, having the highest level of match to the input sentence, the SR system having a predetermined threshold for the level of match, beyond which, the sentence, from the set, that has the highest level of match, is output as a recognized input sentence, is provided. A verbal input sentence is received from the user in the SR system and a weight is assigned for the level of match for each sentences from the plurality according to the content of the input sentence. N sentences having the highest weight from the plurality are selected. If the weight of at least one of the N sentences is higher than the threshold, the sentence having the highest weight is output as the recognized input sentence. If the weight of each selected sentence is lower than the threshold, the weight of each selected sentence is varied according to different predetermined matching criteria. If the varied weight of at least one of the N sentences is higher than the threshold, the sentence having the highest varied weight is output as the recognized input sentence, otherwise providing indication that correspond to unrecognized input sentence.

The input sentence may be recognized by further including the assistance of a human agent that is connected to the SR system. The set of predetermined number of N sentences to be displayed is forwarded to the agent and the input sentence is played to the agent, so as to allow the agent to select a sentence from the set to be output as the recognized input sentence to be output as the recognized input sentence. Alternatively, the input sentence is played to the agent, and the agent is allowed to recognize the input sentence and to type at least a portion of one or more words from the recognized input sentence. If the complete input sentence is typed by the agent, the typed sentence is output as the recognized input sentence. Otherwise, one or more partially typed words are automatically completed and a sentence consisting of completed words is output as the recognized input sentence. Alternatively, the input sentence is played to the agent and the agent is allowed to recognize the input sentence and to recite the recognized input sentence to a voice recognition unit adapted to recognize the voice of the agent according to specific parameters that can be access by the voice recognition unit. A sentence that corresponds to the recited input sentence is output by the voice recognition unit, as the recognized input sentence.

According to one aspect of the invention, the method further comprises recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps: A plurality of input sentences said by a corresponding plurality of users are received and available human agents are allocated. a set of predetermined number of N sentences that are associated with an input sentence, are forwarding to be displayed to each available agent. A different input sentence is played to each available agent and the available agent is allowed to select a sentence from the set to be output as the corresponding recognized input sentence.

According to another aspect of the invention, the method further comprises recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps: A plurality of input sentences said by a corresponding plurality of users are received and available human agents are allocated. The input sentence is played to the agent and the agent is allowed to recognize the input sentence and to type at least a portion of one or more words from the recognized input sentence. If the complete input sentence is typed by the agent, the typed sentence is output as the recognized input sentence. Otherwise, one or more partially typed words are automatically completed and a sentence consisting of completed words is output as the recognized input sentence.

According to another aspect of the invention, the method further comprises recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps: A plurality of input sentences said by a corresponding plurality of users are received and available human agents are allocated. The input sentence is played to the agent and the agent is allowed to recognize the input sentence and to recite the recognized input sentence to a voice recognition unit, adapted to recognize the voice of the agent according to specific parameters that can be access by the voice recognition unit. A sentence that corresponds to the recited input sentence is output by the voice recognition unit, as the recognized input sentence.

The weight of each selected sentence may be varied by evaluating the logic meaning of each selected sentence that consists of objects and the logic relation between them, according to comparisons of the objects and the logic relation between them, to different combinations of predetermined, and essentially similar, objects and the logic relation between them, based on human's common knowledge that is relevant to the subject of the

selected sentences, and by assigning higher weights to one or more selected sentences, each of which having a logic meaning, according to the level of similarity of its logical meaning, to the logical meaning represented by the essentially similar objects and the logic relation between them. If the assigned weight of at least one of the selected sentences is higher than the threshold, the selected sentence having the highest weight is output as the recognized input sentence.

The weight of each selected sentence may also be varied by evaluating each selected sentence, according to the context of the selected sentence with respect to previously recited sentences by the user, with respect to expected objects and/or indirect objects and/or subjects that are essentially related to the content of the previously recited sentences and by assigning higher weights to one or more selected sentences, having closer context relation to previously recited sentences. If the assigned weight of at least one of the selected sentences is higher than the threshold, the selected sentence having the highest weight is output as the recognized input sentence.

The weight of each selected sentence may also be varied by evaluating each selected sentence, according to the context of the selected sentence with respect to expected subsequent state(s) of interaction between the user and the system to which the output of the SR system is input and assigning higher weights to one or more selected sentences, having closer context relation to an expected subsequent state. If the assigned weight of at least one of the selected sentences is higher than the threshold, the selected sentence having the highest weight is output as the recognized input sentence.

The present invention is also directed to an improving speech recognition system, that comprises:

- a) Speech Recognition (SR) unit for receiving a verbal input sentence and outputting a sentence that matches an input sentence of a user, the SR system comprises a plurality of predetermined sentences that are associated with a specific subject, and a set of predetermined number of N sentences, selected from the plurality of predetermined sentences, having the highest level of match to the input sentence, the SR system having a predetermined threshold for the level of match, beyond which, the sentence, from the set, that has the highest level of match, is output as a recognized input sentence; and
- b) processing means for assigning a weight for the level of match for each sentences from the plurality according to the content of the input sentence and for selecting N sentences having the highest weight from the plurality; for outputting the sentence having the highest weight as the recognized input sentence, if the weight of at least one of the N sentences is higher than the threshold; for varying the weight of each selected sentence according to different predetermined matching criteria, if the weight of each selected sentence is lower than the threshold; and for outputting the sentence having the highest varied weight as the recognized input sentence, if the varied weight of at least one of the N sentences is higher than the threshold or providing indication that correspond to unrecognized input sentence.

The system may further comprise a call center that is connected to the SR system and linked to a human agent(s), for recognizing the input sentence using the assistance of the human agent(s). Preferably, the system comprises:

- a) a control unit for receiving a plurality of input sentences said by a corresponding plurality of users; for allocating available human agents; and for forwarding a set of predetermined number of N sentences that are associated with an input sentence, to be displayed to each available agent;

- b) circuitry for playing a different input sentence to each available agent;
and
- c) circuitry for outputting the corresponding recognized input sentence that is selected by the available agent from the set, to be output as the corresponding recognized input sentence.

The system that comprises a call center that is connected to the SR system and linked to a human agent(s), for recognizing the input sentence using the assistance of the human agent(s), may further comprise:

- a) a control unit for receiving a plurality of input sentences said by a corresponding plurality of users; and for allocating available human agents;
- b) circuitry for playing a different input sentence to each available agent;
- c) input means for typing at least a portion of one or more words from the recognized by the available agent; and
- d) circuitry for outputting the typed sentence as the recognized input sentence and/or computerized means for automatically completing one or more partially typed words before outputting the partially typed sentence.

The system that comprises a call center that is connected to the SR system and linked to a human agent(s), for recognizing the input sentence using the assistance of the human agent(s), may further comprise:

- a) a control unit for receiving a plurality of input sentences said by a corresponding plurality of users; and for allocating available human agents;
- b) circuitry for playing a different input sentence to each available agent; and
- c) voice recognition unit for recognizing an input sentence recited by the available agent according to specific parameters that can be

access by the voice recognition unit and for outputting a sentence corresponding to the recited input sentence, as the recognized input sentence.

Brief Description of the Drawings

The above and other characteristics and advantages of the invention will be better understood through the following illustrative and non-limitative detailed description of preferred embodiments thereof, with reference to the appended drawings, wherein:

- Fig. 1 schematically illustrates a conventional voice recognition system; and
- Fig. 2 schematically illustrates an enhanced voice recognition system, according to a preferred embodiment of the invention.

Detailed Description of Preferred Embodiments

Fig. 1 schematically illustrates a voice recognition system 100 according to a prior art. ASR 102 receives as its input sentences from a user, which could be sent, for example, by phone 101. ASR 102 tries to guess the sentence of the user, according to predetermined sentences (phrased by using given words and grammar rules for making the sentences) and by the threshold limit. At the end of the recognition process, ASR 102 provides N-best sentences with their weight. If there is any sentence, among the N-best sentence, which passed the threshold and has the highest weight, this sentence will be provided as an output of ASR 102, and will represent the sentence that probably said by the user for further processing. If there is no such sentence, then ASR 102 failed in recognizing what was the sentence that has been said by the user.

Fig. 2 schematically illustrates an enhanced voice recognition system 200. ASR 102 receives a sentence as an input from a user, which could be sent, for example, by phone 101. ASR 102 tries to guess the sentence received

from the user, according to predetermined sentences (phrased by using given words and grammar rules for making the sentences) and to the threshold limit. At the end of recognition process, ASR 102 provides N-best sentences with their corresponding weight. If there is any sentence, among the N-best sentence, which passed the threshold and has the highest weight, this sentence will be provided as an output of ASR 102 and will represent the sentence that was probably said by the user for further processing. If there is no such sentence, then ASR 102 failed in recognizing what was actually the sentence that has been said by the user.

According to a preferred embodiment of the invention, in case where ASR 102 failed to recognize the sentence of the user, the N-best sentences that hasn't passed the threshold are transferred to a Common Sense module 201 for further processing. Common Sense module 201 increases the weight of each sentence among the N-best sentences that has a logical meaning in reality, and/or decrease the weight of each sentence among the N-best sentences that has less logical meaning in reality, even though it grammatically correct. Taking for example, the sentence "the books are swimming in the air", even though it grammatically correct, it has no logical meaning in reality. Common Sense module 201 uses ontology component (not shown in Fig. 2), which is a computer representation of a specific human vision of the actual reality, among several different human visions of said reality, in order to obtain logical and meaningful relation between the words of each N-best sentence. According to a preferred embodiment of the invention, the ontology component may be implemented as a predetermined database that comprises objects (usually nouns) and the logical relation between them. After the Common Sense module 201 has completed to check all the N-best sentences, and changed their weights as described hereinabove, there may be one or more sentences that will pass the threshold of the ASR. The sentence that has the highest weight and passes the threshold will represent the sentence

that said by the user for further processing by the system (not shown), to which the ASR 102 is attached. At that point the recognition process of the system 200 is completed and paused, until a new sentence will be entered to system 200.

If non-of the N-best sentences have passed the threshold after the processing of the Common Sense module 201, then according to the preferred embodiment of the invention, the N-best sentences with their new weights (determined by the Common Sense module 201) will be transferred to a Context Handling module 202. Context Handling module 202 attempts to increase the weight of each sentence among the N-best sentences with words that represent objects, indirect objects or subjects that has exist in the context of the previous sentences said during the conversation, and recognized confidently by the ASR. In addition, the weight of each sentence among the N-best sentences that has no context with the previous sentences of the conversation, may be decreased. Context Handling module 202 is used to track the conversation in order to obtain the user's intention at any time. The Context Handling module 202 may store subjects, objects and indirect objects that were mentioned directly or indirectly during interaction with the user, and may be related to items stored in an accessible database. For example, during a conversation, a user may use the term 'it' in a sentence instead of a noun used in an earliest sentence. In another example, a user has mention in earlier sentence a name of a movie, and at the current sentence there is a name of an actor. After the Context Handling module 202 completed to check all the N-best sentences, and changed their weights accordingly (as described hereinabove), there may be one or more sentences that will pass the threshold of the ASR 102. The sentence that has the highest weight and passes the threshold, will represent the sentence that said by the user for further processing by the system (not shown), to which the ASR 102 is

attached. At that point the recognition process of the system 200 is completed and paused, until a new sentence will be entered to system 200.

If non-of the N-best sentences have passed the threshold after the Context Handling module 202, then according to the preferred embodiment of the invention, the N-best sentences with their new weights will be transferred to a Flow Handling module 203. Flow Handling module 203 is based on a principle of state-machine (i.e., system 200 "knows" which are the possible next steps it can make, according to the previous status of the conversation). Flow Handling module 203 increases the weight of a sentence, among the N-best, that will allow the system to move from the current state of the conversation, to the next possible state. The weight of the sentence that won't fulfill this state-condition will be decreased. After the Flow Handling module 203 has completed checking all the N-best sentences, and changed their weights as described hereinabove, there may be one or more sentences that will pass the threshold of the ASR. The sentence that has the highest weight and passed the threshold, will represent the sentence that said by the user for further processing by the system (not shown), to which the ASR 102 is attached. At that point the recognition process of the system 200 is completed and paused, until a new sentence will be entered to system 200.

It is important to mention that each of the modules 201, 202 and 203 is independent and may change the weights of the N-best sentences according to its criteria. One of the modules may increase the weight of a specific sentence and the other module may decrease the weight of this specific sentence.

However, if non-of the sentences among the N-best sentences has passed the threshold of ASR 102, after the last module 203, than the N-best sentences are transferred to a Human Assistance Manager (HAS

Manager) 204, which directs the input information to an available agent of a Call Center 205 (i.e., to human assistance that intends and minimally qualified only to simple voice recognition tasks). HAS manager 204 receives the unrecognized input sentence of the user and the N-best possibilities of recognition for this input sentence. In this way, the Agent of a Call Center 205 does the recognition of the user's sentence in a very short time, since the Agent of a Call Center 205 hears only the user sentence that has not been successfully recognized (out of the whole conversation time period). Agent of a Call Center 205 hears only the unrecognized sentence of the user and recognizes it by choosing on of the following three alternatives, which are selecting, typing or speaking:

- The N-best sentences are introduced to an Agent of a Call Center 205 and according to block 206, he selects the sentence with closest match to the user's sentence that was heard.
- Agent of a Call Center 205 recognize according to block 207 the sentence of the user that was heard, and he directly types said sentence. In order to save human resources (i.e., to shorten the typing time), parts of the sentence that is typed by the Agent of a Call Center 205 are completed by executing a Completing Application 209 that completes words that are partially typed by the Agent of a Call Center 205. The completion is carried out using complete words that are stored in a database and match the characters that are partially typed.
- Agent of a Call Center 205 hears the user sentence and according to block 208 he recites said sentence clearly (using his own voice) to a Voice Recognition (VR) unit 210. This VR unit 210 is "trained" for the task of recognizing this specific voice profile of the Agent of a Call Center 205 (the "training" is carried out by receiving the Agent's voice several times in different variations in his voice in lab conditions). In such case, the recognition accuracy is substantially improved.

After Agent of a Call Center 205 selects one of the three options 206, 207, or 208, the recognized sentence is delivered to further processing, by the system (not shown), to which the ASR 102 is attached. At that point the recognition process of the system 200 is completed and paused, until a new sentence will be entered to system 200.

In order to reduce time of recognition by Agent of a Call Center 205, he receives only the unrecognized sentence from a complete session with the user, instead of listening to the complete conversation with the user.

Due to fact that human resources are exploited by system 200, it is desired to reduce the time during which human resource are exploited during each session. In conventional conversation, the relation between a user and the HAS is 1:1, i.e., human assistance is required during a complete session with a user. This relation is reduced, so as to allow the Agent of a Call Center 205 to simultaneously handle several sessions.

A typical session with a user consists of the following time periods:

- Entry/Exit time, small grammar talk	X seconds
- Question posed	X seconds
- Computer processing ("thinking") time	0.5X seconds
- Computer response time	2X seconds
- Human processing ("thinking") time	X seconds
- HAS processing time	1.5X seconds
<hr/>	
Total session time	7X seconds

The Entry/Exit time reflects the time required for greetings, such as "Hello"/"Bye", respectively. This section of the conversation is fixed and does not require HAS intervention.

Small grammar talks reflects the confirmation words, such as "Tell me Yes or No". Such grammar has high probability to be recognized by automate measures, such as ASR, due to its small size (for example, "Yes" or "No"). This section of the conversation does not require HAS intervention.

Question posed reflects the time required for a user to introduce a question/request (this is normally determines the time period "X", as the relative time period to estimate all sections of the conversation). This section of the conversation does not require HAS intervention.

Computer processing ("thinking") time reflects the time required to perform the operation of modules 201 to 203, ASR 102 and further processing by the system (not shown), to which the ASR 102 is attached. This section of the conversation does not require HAS intervention.

Computer response time reflects the time required for play the response to the user. In this section the computer is "Speaking", while the user listen. This section of the conversation does not require HAS intervention.

Human processing ("thinking") time reflects the time required for the user to process and understands the computer response, and the time to think about his next question/request. This section of the conversation does not require HAS intervention.

HAS processing time reflects the time required for the HAS to convert an unrecognized sentence, to a string, in one of the three ways that described hereinabove.

If, for example, the average probability of a conventional ASR system to provide a correct recognition is 65%, than in order to increase the

probability to be close to 100%, additional improvement of approximately 25% is required from HAS 220 (provided that modules 201 to 203 can improved the probability by 10%). Therefore, the average conversation segment length is:

$$0.25*7X + (0.65+0.1)*5.5X = 5.88X$$

7X - is total conversation segment, containing HAS intervention.

5.5X - is total conversation segment, which does not containing HAS intervention.

Since assistance from HAS 220 is required once for four conversation segments (i.e., 0.25), the 1.5X seconds of HAS assistance spend once during $4*5.88X$ time period => ratio of 1:15.68. Therefore, the HAS 220 assistance (Human Resource) is required only one second of every 15.68 seconds of speech. This lead us, in call center supporting hundreds to thousands of concurrent users, i.e., where the principle of "big numbers" will work, which require about 100% recognition, we will need 1 Human Resource per every 15 concurrent users.

The above examples and description have of course been provided only for the purpose of illustration, and are not intended to limit the invention in any way. As will be appreciated by the skilled person, the invention can be carried out in a great variety of ways, employing more than one technique from those described above, all without exceeding the scope of the invention.

CLAIMS

1. A method for improving speech recognition, comprising:
 - a) providing a Speech Recognition (SR) system for outputting a sentence that matches an input sentence of a user, said SR system comprises a plurality of predetermined sentences that are associated with a specific subject, and a set of predetermined number of N sentences, selected from said plurality of predetermined sentences, having the highest level of match to said input sentence, said SR system having a predetermined threshold for said level of match, beyond which, the sentence, from said set, that has the highest level of match, is output as a recognized input sentence;
 - b) receiving a verbal input sentence from said user in said SR system and assigning a weight for the level of match for each sentences from said plurality according to the content of said input sentence;
 - c) selecting N sentences having the highest weight from said plurality;
 - d) if the weight of at least one of said N sentences is higher than said threshold, outputting the sentence having the highest weight as the recognized input sentence;
 - e) if the weight of each selected sentence is lower than said threshold, varying the weight of each selected sentence according to different predetermined matching criteria; and
 - f) if the varied weight of at least one of said N sentences is higher than said threshold, outputting the sentence having the highest varied weight as the recognized input sentence, otherwise providing indication that correspond to unrecognized input sentence.
2. A method according to claim 1, further comprising recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps:

- a) forwarding the set of predetermined number of N sentences to be displayed to said agent; and
 - b) playing said input sentence to said agent and allowing said agent to select a sentence from said set to be output as the recognized input sentence and outputting said recognized input sentence.
3. A method according to claim 1, further comprising recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps:
- a) playing said input sentence to said agent and allowing said agent to recognize said input sentence and to type at least a portion of one or more words from the recognized input sentence;
 - b) if the complete input sentence is typed by said agent, outputting the typed sentence as the recognized input sentence; otherwise
 - c) automatically completing one or more partially typed words and outputting a sentence consisting of completed words as the recognized input sentence.
4. A method according to claim 1, further comprising recognizing the input sentence using the assistance of a human agent that is connected to the SR system; by performing the following steps:
- a) playing said input sentence to said agent and allowing said agent to recognize said input sentence and to recite the recognized input sentence to a voice recognition unit adapted to recognize the voice of said agent according to specific parameters that can be access by said voice recognition unit; and
 - b) outputting, by said voice recognition unit, a sentence corresponding to the recited input sentence as the recognized input sentence.

5. A method according to claim 1, further comprising recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps:
 - a) receiving a plurality of input sentences said by a corresponding plurality of users;
 - b) allocating available human agents;
 - c) forwarding a set of predetermined number of N sentences that are associated with an input sentence, to be displayed to each available agent; and
 - d) playing a different input sentence to each available agent and allowing said available agent to select a sentence from said set to be output as the corresponding recognized input sentence and outputting said corresponding recognized input sentence.
6. A method according to claim 1, further comprising recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps:
 - a) receiving a plurality of input sentences said by a corresponding plurality of users;
 - b) allocating available human agents;
 - c) playing said input sentence to said agent and allowing said agent to recognize said input sentence and to type at least a portion of one or more words from the recognized input sentence;
 - d) if the complete input sentence is typed by said agent, outputting the typed sentence as the recognized input sentence; otherwise
 - e) automatically completing one or more partially typed words and outputting a sentence consisting of completed words as the recognized input sentence.

7. A method according to claim 1, further comprising recognizing the input sentence using the assistance of a human agent that is connected to the SR system, by performing the following steps:
- a) receiving a plurality of input sentences said by a corresponding plurality of users;
 - b) allocating available human agents;
 - c) playing said input sentence to said agent and allowing said agent to recognize said input sentence and to recite the recognized input sentence to a voice recognition unit adapted to recognize the voice of said agent according to specific parameters that can be access by said voice recognition unit; and
 - d) outputting, by said voice recognition unit, a sentence corresponding to the recited input sentence as the recognized input sentence.
8. A method according to claim 1, wherein the weight of each selected sentence is varied by performing the following steps:
- a) evaluating the logic meaning of each selected sentence that consists of objects and the logic relation between them, according to comparisons of said objects and the logic relation between them, to different combinations of predetermined, and essentially similar, objects and the logic relation between them, based on human's common knowledge that is relevant to the subject of said selected sentences;
 - b) assigning higher weights to one or more selected sentences, each of which having a logic meaning, according to the level of similarity of its logical meaning, to the logical meaning represented by said essentially similar objects and the logic relation between them; and
 - c) if the assigned weight of at least one of said selected sentences is higher than the threshold, outputting the selected sentence having the highest weight as the recognized input sentence.

9. A method according to claim 1, wherein the weight of each selected sentence is varied by performing the following steps:
- a) evaluating each selected sentence, according to the context of said selected sentence with respect to previously recited sentences by the user, with respect to expected objects and/or indirect objects and/or subjects that are essentially related to the content of said previously recited sentences;
 - b) assigning higher weights to one or more selected sentences, having closer context relation to previously recited sentences; and
 - c) if the assigned weight of at least one of said selected sentences is higher than the threshold, outputting the selected sentence having the highest weight as the recognized input sentence.
10. A method according to claim 1, wherein the weight of each selected sentence is varied by performing the following steps:
- a) evaluating each selected sentence, according to the context of said selected sentence with respect to expected subsequent state(s) of interaction between the user and the system to which the output of the SR system is input;
 - b) assigning higher weights to one or more selected sentences, having closer context relation to an expected subsequent state; and
 - c) if the assigned weight of at least one of said selected sentences is higher than the threshold, outputting the selected sentence having the highest weight as the recognized input sentence.
11. An improving speech recognition system, comprising:
- a) Speech Recognition (SR) unit for receiving a verbal input sentence and outputting a sentence that matches an input sentence of a user, said SR system comprises a plurality of predetermined sentences that are associated with a specific subject, and a set of predetermined number of N sentences, selected from said plurality

of predetermined sentences, having the highest level of match to said input sentence, said SR system having a predetermined threshold for said level of match, beyond which, the sentence, from said set, that has the highest level of match, is output as a recognized input sentence;

- b) processing means for assigning a weight for the level of match for each sentences from said plurality according to the content of said input sentence and for selecting N sentences having the highest weight from said plurality; for outputting the sentence having the highest weight as the recognized input sentence, if the weight of at least one of said N sentences is higher than said threshold; for varying the weight of each selected sentence according to different predetermined matching criteria, if the weight of each selected sentence is lower than said threshold; and for outputting the sentence having the highest varied weight as the recognized input sentence, if the varied weight of at least one of said N sentences is higher than said threshold or providing indication that correspond to unrecognized input sentence.

12. A system according to claim 11, further comprising a call center that is connected to the SR system and linked to a human agent(s), for recognizing the input sentence using the assistance of said human agent(s), comprising:

- a) a control unit for receiving a plurality of input sentences said by a corresponding plurality of users; for allocating available human agents; and for forwarding a set of predetermined number of N sentences that are associated with an input sentence, to be displayed to each available agent; and
- b) circuitry for playing a different input sentence to each available agent; and

- c) circuitry for outputting said corresponding recognized input sentence that is selected by said available agent from said set, to be output as the corresponding recognized input sentence.
13. A system according to claim 11, further comprising a call center that is connected to the SR system and linked to a human agent(s), for recognizing the input sentence using the assistance of said human agent(s), comprising:
- a) a control unit for receiving a plurality of input sentences said by a corresponding plurality of users; and for allocating available human agents;
 - b) circuitry for playing a different input sentence to each available agent;
 - c) input means for typing at least a portion of one or more words from the recognized by said available agent; and
 - d) circuitry for outputting the typed sentence as the recognized input sentence and/or computerized means for automatically completing one or more partially typed words before outputting the partially typed sentence.
14. A system according to claim 11, further comprising a call center that is connected to the SR system and linked to a human agent(s), for recognizing the input sentence using the assistance of said human agent(s), comprising:
- a) a control unit for receiving a plurality of input sentences said by a corresponding plurality of users; and for allocating available human agents;
 - b) circuitry for playing a different input sentence to each available agent; and
 - c) voice recognition unit for recognizing an input sentence recited by said available agent according to specific parameters that can be

access by said voice recognition unit and for outputting a sentence corresponding to the recited input sentence, as the recognized input sentence.

15. A method for improving speech recognition, substantially as described and illustrated.
16. A system for improving speech recognition, substantially as described and illustrated.

1/3

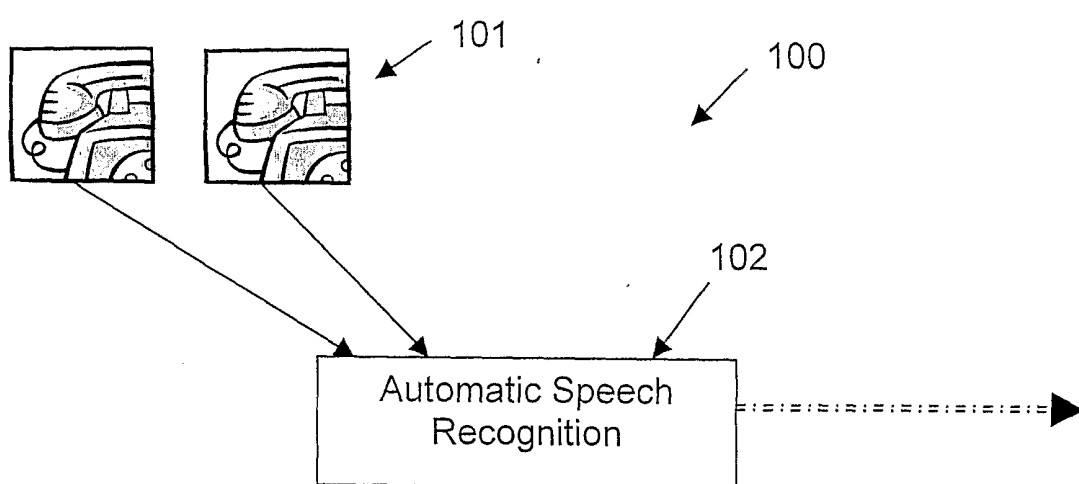


Fig. 1

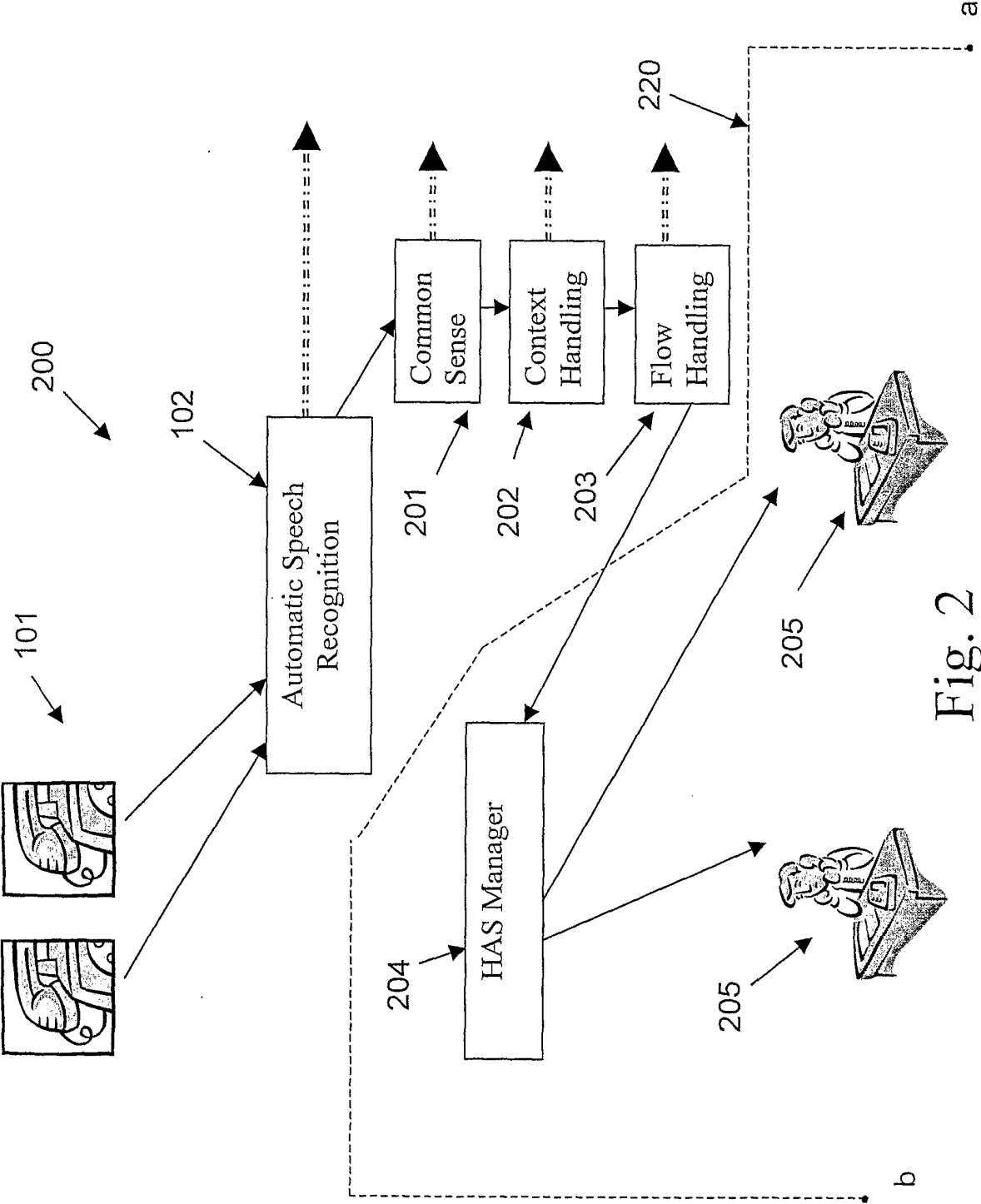


Fig. 2

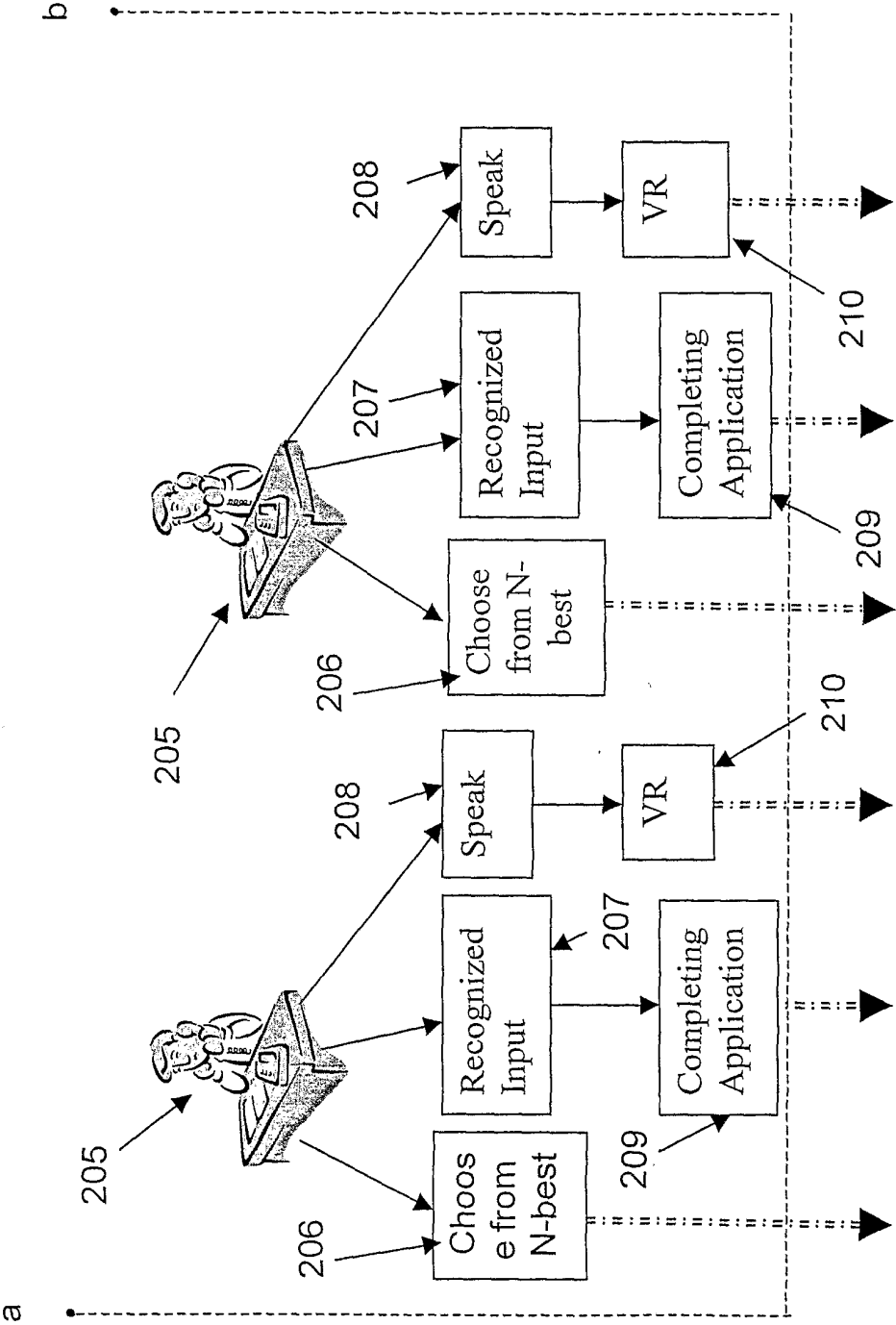


Fig. 2(cont.)