

Text-processing arrangement.

The invention relates to a text-processing arrangement as defined in the introductory part of the main Claim. Such a text-processing arrangement is described in US-PS 3,770,892, but this document does not describe how to correct words which have not been detected correctly by the speech-recognition device. Since errors in the detection of words occur comparatively often in existing text-processing arrangements of this type such a testing or correction facility is of special importance.

It is an object of the invention to construct a text-processing arrangement of the type defined in the opening sentence in such a way that the detected words can be checked and erroneously detected words can be corrected simply, rapidly and reliably. To this end the arrangement has the characteristic features defined in the characterising part of the main Claim. Since the audio signals corresponding to a text are stored additionally in a first further memory and, moreover, in correlation with these stored sound signals the start marks and end marks, which are related in time to the words of the text detected by the speech recognition device and stored in a word memory, are also stored in a second further memory, it is simply possible at any time to allocate the associated sound signals to each detected word stored in the word memory or, conversely, to allocate to the corresponding detected word to each portion of the sound signal. Thus, the user of the arrangement can, for example, acoustically monitor the original sound signal which is associated with a word to be checked or corrected and which is highlighted on the display device by means of the computer, so that the user can correctly identify the relevant original word and, if necessary, enter said word in corrected form into the arrangement *via* the keyboard. Another possibility is to carry out a continuous check of the detected words of the text as displayed by the display device in that the sound signals stored in the first further memory are continually reproduced acoustically, the associated detected word being highlighted simultaneously on the display device by means of the computer, so that the instantaneously detected word is highlighted continually and the user of the arrangement at the same time hears

the associated original sound signal, thereby enabling words requiring correction to be recognized simply and accurately by the user.

An arrangement in accordance with the invention, which in addition has the features defined in the opening part of the 2nd Claim preferably has the
5 characteristic features defined in the characterising part of this Claim. In this way words which are initially detected incorrectly and which are subsequently corrected are also included in a learning process of the arrangement, so that during use of the arrangement the error rate of incorrectly identified words decreases continually and consequently the entire text-processing arrangement operates with an increasing reliability.

10 An arrangement in accordance with the invention may also have the characteristic features defined in Claim 3. In this way a learning process of the arrangement allows not only for individual corrected words but also for the entire text which has already been corrected, so that during use of the arrangement not only the error rate of incorrectly detected words is constantly reduced but the arrangement is
15 also rapidly adapted to the characteristic pronunciation of a specific speaker or to specific text sequences frequently used by a specific speaker.

The invention will now be described in more detail reference to the drawing, which shows an exemplary embodiment to which the invention is not limited. The Figure of the drawing shows a block diagram of a text-processing arrangement.

20 In the block diagram shown in the Figure a speech-recognition device bears the reference numeral 1 and the sound signals corresponding to a text are applied in the form of electric signals from an input 3 of the text-processing arrangement to an input 2 of said device, said sound signals being obtained, for example, directly from a microphone or being reproduced by means of a dictation machine employed as a
25 temporary store. These electric signals may be either analog signals or digital signals, which last-mentioned signals may be in encoded form, depending on the type of speech-recognition device used, a variety of such devices suitable for use in the present case being known. The speech-recognition device 1 is controlled by means of a computer 4, which also controls the entire text-processing arrangement. From the input 2 of the
30 speech recognition device 1 the electric signal is applied in customary manner to an acoustic analyzer 5, which determines salient features which are characteristic of this signal, for example energy-concentration poles within the frequency spectrum, which are subsequently stored in a feature memory 6. This feature memory 6 provides an

input to the actual word-recognition device 7 of the speech-recognition device 1, which is also connected to a series of memories enabling individual words to be identified by the device. In the customary manner these memories comprise a speech-model memory 8 for the storage of probable word sequences enunciated by a specific speaker or several speakers, a phonetics memory 9 for the storage of sound compositions, which may be speaker-related, for the words which can be recognized by the speech recognition arrangement, and a pronunciation-feature memory 10 for the storage of parameters relating to the pronunciation characteristic of a specific speaker or several speakers. In the present case the speech-recognition device 1 further comprises a learning device 11, which cooperates directly with the word-recognition device 7 and the feature memory 6 and which also has access to the above-mentioned memories 8, 9 and 10. In the course of a learning process the features derived from a sound signal enunciated by a specific speaker of a word by means of the speech-recognition device 1 and stored in the feature memory 6 are analyzed by the learning device 11 by comparison with speech-related parameters which are stored, for example, in the memories 8, 9 and 10 and which are applied to the learning device, and the word which is associated with the relevant sound signal and which is also applied to the learning device, for example by the computer 4, which features are stored in the memories 8, 9 and 10 of the speech-recognition device as speaker-related parameters, so that they are henceforth available to the word-recognition device 7.

Said means enable the individual words of the complete text contained in the sound signals to be recognized, which words are subsequently stored in a word memory 12 as digital signals, in which memory they are then available for further processing in a processing device (4, 13, 14, 17) and can be applied to a display device 13 or a printer 14 via the computer 4. In this way the text represented by the sound signal can be displayed directly on the display device 13 as a text made up of individual words, where it is available for checking or correction, which is necessary because such speech-recognition devices generally exhibit a specific error rate, which implies that individual words or word groups are not identified correctly and therefore appear incorrectly on the display.

In order to facilitate the correction of erroneously detected words the present text-processing arrangement utilizes further steps, which will be explained hereinafter.

In particular, the sound signals applied to the input 3 of the text-processing arrangement as electric signals are stored in a first further memory 15. For the storage of the sound signals in the first further memory 15 various possibilities are available, depending on the construction of this further memory. For example, such a first further memory may be an audio-tape apparatus, on which the audio signals are stored as analog signals. However, such a first further memory may alternatively be a digital memory, to which the audio signals are applied after digitization or encoding, such memories having the advantage of a more rapid access to the memory contents. Said memory may be, for example, a disk storage or a solid-state memory. Moreover, the memory of a central dictating arrangement may also be used as such a first further memory. Alternatively, the input signal to be applied to the input 2 of the speech recognition device 1 itself may be provided by an output of such first further memory.

Moreover, start marks and end marks are allocated to the individual words detected by the speech-recognition device and stored in the word memory 12, which marks are in a specific time relationship with the sound signals stored in the first further memory 15. In this way every detected word is bounded by a start mark and an end mark, and the end mark of a word may at the same time form the start mark of the next word. In this respect it is essential that in generating these marks it is guaranteed that they are in a specific time-relationship with the sound signals stored in the first further memory 15. For this purpose a time-counting process may be carried out during the speech-recognition process, for example by means of the clock signal of the computer 4, to consecutively define the individual time intervals occurring between individual word detections: The start and end marks of the individually detected words thus generated are stored in a second further memory 16, which is connected to the word-recognition device 7, access to this second further memory 16 being provided via the computer 4. The computer 4 also has access to the first further memory 15, in which the audio signals are stored. By means of the start mark and end mark of a detected word stored in the second further memory 16 the computer 4 can thus establish a correlation with the sound signal associated with this detected word and stored in the first further memory 15, enabling a detected word displayed by means of the display device 13 to be highlighted, for example with a cursor or by underscoring, under control of the computer 4 and at the same time the corresponding audio signal to be reproduced acoustically.

Thus, the text formed by the words displayed by means of the display device 13 can be checked or corrected simply, rapidly and reliably. By means of a keyboard 17, which cooperates with the computer 4 and which serves for the entry of instructions and data, a highlighted word displayed on the display device 13 can be corrected, if necessary. For example, within the complete text formed by the words displayed on the display device a single word of this text can be highlighted under control of the keyboard 17 and the computer 4, after which by means of the computer 4 the start mark and end mark corresponding to the highlighted word and stored in the second further memory 16 are detected and, depending upon these marks, the corresponding audio signals stored in the first further memory 15 can be read out and can be reproduced acoustically, for example by means of a loudspeaker 18, which reproduction may be non-recurrent or recurrent. In this way the user of the arrangement has a direct possibility of monitoring the audio signal corresponding to the relevant highlighted word in order to verify whether the relevant word has been detected correctly or erroneously by means of the speech-recognition device 1. In the case of an erroneously detected word the user of the arrangement can replace the erroneously detected word by the correct word by means of the keyboard 17. However, it is also possible to perform for example a continuous check of the text displayed by means of the display device 13 in such a way that the sound signals stored in the first further memory 15 are continuously reproduced acoustically, while at the same time the corresponding detected words are highlighted on the display device 13 by means of the computer 4 in conjunction with the start and end marks of every detected word as stored in the second further memory 16. Thus, the user of the arrangement continuously hears the spoken text and at the same time his attention is drawn to the word detected by the speech-recognition device which is highlighted in the displayed text. It is evident that as a result of these different possibilities of using the described text-processing arrangement a text can be checked and corrected simply and reliably, after which the finished correct text can be printed by means of the printer 14.

It is found to be advantageous if the corrected version of a detected word which has been highlighted on the display device 13 and corrected by means of the keyboard 17 together with the features stored in the feature memory 6, which features have been derived from the sound signal of the original word by means of the speech-recognition device 1, is applied to the learning device 11 for analysis, the speaker-

related parameters which are newly derived in the course of this analysis being also stored in the memories 8, 9 and 10 of the speech-recognition device 1. In the present case it is assumed that all the characteristic features derived from the sound signals of the entire text by means of the speech-recognition device 1 are continuously available for such learning processes in the feature memory 6 and the features corresponding to the sound signal of a word can always be read out, which requires a feature memory 6 having a comparatively large storage capacity. Another possibility is to utilize the sound signals stored in the first further memory 15 for such learning processes and to re-detect the relevant features each time by means of the speech-recognition device 1 and store them again in the feature memory 6, so that the storage capacity of the feature memory 6 can then be smaller. For this purpose it is possible, for example, to connect the input 2 of the speech-recognition device 1 to an output of the first further memory 15 instead of to the input 3 of the arrangement by means of a switch 19. The computer 4 then again provides access to every sound signal portion stored in the first further memory 15 and corresponding to a relevant word detected by the speech-recognition device. In this way the learning device 11 modifies or updates the speaker-related parameters stored in the memories 8, 9 and 10 of the speech-recognition device 1 in such a manner that the speech-recognition device 1 performs a further learning process relating to the relevant speaker to ensure that if in future a sound signal corresponding to such a word occurs again the probability of said word being correctly detected at once by the speech recognition device 1 increases, so that the error rate of such a text-processing arrangement is thus reduced continually.

It is also found to be advantageous if after all the words which have been detected erroneously by the speech-recognition device 1 have been corrected by means of the keyboard 17 the words forming the complete text together with the features derived from the sound signals of the original words by means of the speech-recognition device and stored in the feature memory 6 are applied to the learning device 11 for analysis, the newly derived speaker-related parameters obtained in the course of this analysis being also stored in the memories 8, 9 and 10 of the speech-recognition device. In this way the entire text already corrected is utilized in a learning process of the arrangement, so that both the error rate of erroneously detected words is reduced and a rapid adaptation of the arrangement to the characteristic pronunciation of a specific

speaker and to specific phrases frequently used by the relevant speaker is achieved, resulting in a substantial improvement of the performance of such an arrangement.

As will be apparent from the foregoing a variety of modifications of the described exemplary embodiment are possible without departing from the scope of the invention. This applies in particular to the manner in which the sound signals are stored in the first further memory, the manner of generating and analyzing the start mark and the end mark which bound a word detected by the speech-recognition device, and the operation of the actual speech-recognition device. It is obvious that such a text-processing arrangement can also perform other functions which are customary in such arrangements, such as for example text merging or the like.

Claims:

1. A text-processing arrangement, comprising
- a speech-recognition device (1) for receiving sound signals corresponding to a text and for detecting the words contained in the sound signals,
 - a word memory (12) for storing the detected words in the form of digital signals,
 - 5 - a processing device (4, 13, 14, 17) connected to the word memory (12) and adapted to receive the detected words in the form of digital signals, which processing device comprises a display device (13) for reproducing the detected words as text and a correction device (17) for the correction of erroneously detected words, characterised in that there is provided a first further memory (15) for storing the sound signals, in that
 - 10 the processing device (4, 13, 14, 17) comprises a computer (4) which cooperates with a keyboard (17) and which is adapted to control the speech-recognition device (1), in that there are provided means (1) for allocating start marks and end marks to the words stored in the word memory (12), which start marks and end marks are related in time to the sound signals stored in the first further memory (15), in that there is provided a
 - 15 second further memory (16) for storing the start marks and end marks, in that the computer (4) is further adapted to establish a correlation between a word stored in the word memory (12) and the sound signal stored in the first further memory (15) and corresponding to the word *via* the start mark and the end mark corresponding to this word and stored in the second further memory (16), in that the computer (4) is further
 - 20 adapted to highlight at least one detected word displayed on the display device (13) and to control means (15, 18) for the simultaneous acoustic reproduction of the sound signal associated with the one displayed and highlighted word, and in that the keyboard (17) is adapted to load the computer (4) with data for, if necessary, correcting the word displayed and highlighted on the display device (13).
- 25 2. An arrangement as claimed in Claim 1, comprising a learning device (11) for cooperation with the speech-recognition device (1), the speech-recognition device (1) being adapted to derive salient characteristic features from a sound signal which corresponds to at least one word enunciated by a speaker and to store said features in



the form of speaker-related parameters utilized by the speech-recognition device (1) in the process of speech recognition, and the learning device (11) being adapted to receive the features and to receive speech-related parameters and the at least one word corresponding to the sound signal enunciated by the speaker and stored in the word memory (12) and to analyse the received features in comparison with the received speech-related parameters and the received word, characterised in that the learning device (11) is adapted to receive and analyse one word displayed on the display device (13) and corrected by means of the keyboard to form an original correct word from an erroneously detected word, to receive and analyse the features derived from the sound signal corresponding to the original word by means of the speech-recognition device (1), and to derive new speaker-related parameters in the process of the analysis of the corrected word and of the features derived by means of the speech-recognition device, and in that the speech-recognition device (1) is adapted to store the new speaker-related parameters.

3. An arrangement as claimed in Claim 1, comprising a learning device (11) for cooperation with the speech-recognition device (1), the speech-recognition device (1) being adapted to derive salient characteristic features from a sound signal which corresponds to at least one word enunciated by a speaker and to store said features in the form of speaker-related parameters utilized by the speech-recognition device (1) in the process of speech recognition, and the learning device (11) being adapted to receive the features and to receive speech-related parameters and the at least one word corresponding to the sound signal enunciated by the speaker and stored in the word memory (12) and to analyse the received features in comparison with the received speech-related parameters and the received word, characterised in that the learning device (11) is adapted to receive and analyse all the words forming a complete text and displayed by the display device (13), of which the words detected erroneously by the speech-recognition device (1) have been corrected by means of the keyboard (17) to form the original words, and to receive and analyse the features derived from the sound signals corresponding to the original correct words of the complete text by means of the speech-recognition device (1), and to derive new speaker-related parameters in the process of the analysis of the partly corrected words forming the entire text and of the features derived by means of the speech-recognition device (1), and in that the speech-recognition device (1) is adapted to store the new speaker-related parameters.

