EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent: 06.04.2005 Bulletin 2005/14

(21) Application number: 00123333.7

(22) Date of filing: 27.10.2000

(54) Normalizing voice pitch for voice recognition

Normalisierung der Grundfrequenz zur Spracherkennung

Normalisation du ton de la voix pour la reconnaissance de la parole

(30) Priority: 29.10.1999 JP 30929099

(43) Date of publication of application: 02.05.2001 Bulletin 2001/18

(73) Proprietor: MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.
Kadoma-shi, Osaka 571-8501 (JP)

(72) Inventors:
• Oda, Mikio Yawata-shi, Kyoto-fu (JP)
• Kawane, Tomoe Nishinomiya-shi, Hyogo-ken (JP)

(74) Representative: Lang, Johannes, Dipl.-Ing. et al Bardehle Pagenberg Dost Altenburg Geissler Isenbruck, Postfach 86 06 20 81633 München (DE)

References cited:

US-A- 5 694 520
US-A- 5 966 687

• HIERONYMUS J L: "FORMANT NORMALISATION FOR SPEECH RECOGNITION AND VOWEL STUDIES" SPEECH COMMUNICATION,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 10, no. 5 / 06, 1 December 1991 (1991-12-01), pages 471-478, XP000271195 ISSN: 0167-6393


Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).
Description

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to voice recognition devices capable of recognizing human voice no matter who is the speaker, e.g., low-pitched man, or high-pitched woman or child, and more specifically, to a device for normalizing voice pitch on the basis of a previously-provided sample voice pitch.

Description of the Background Art

[0002] Recently, with the progression of digital signal processing technology and LSI of higher performance capabilities and lower price, voice recognition technology became popular with consumer electronic products. The voice recognition technology also improves such products in operability. Such voice recognition device principally works to recognize human voice by converting an incoming command voice into a digital voice signal, and then referring to a voice dictionary for sample voice data previously-prepared for comparison. Therefore, for easy comparison, the voice recognition device often requests a user to produce a sound for commanding in a specific manner, or to register the user's voice in advance, for example.

[0003] The issue herein is, specifying a user in the voice recognition device equipped in the consumer electric product badly impairs its usability and thus product value. To get around such problem, the voice recognition device is expected to recognize human voice varied in pitch and speed, no matter who is the speaker. However, as already described, the conventional voice recognition device refers to the voice dictionary for comparison with an incoming command voice. Therefore, if the incoming command voice is differed in pitch or speed to a large extent from the sample in the voice dictionary, the voice recognition device fails to correctly perform voice recognition, see "Formant normalisation for speech recognition and vowel studies" from J.L. Hironymus in Speech Communication 10, pp. 471-478, 1991.


[0005] A sound, or voice produced by a user is taken into the voice input part 111, and is captured as a command voice thereby. The captured command voice is A/D converted into a digital voice signal. The voice speed calculation part 112 receives thus produced digital voice signal, and based thereon, calculates the user's voice speed. The voice speed change rate determination part 113 compares thus calculated voice speed with a reference voice speed, and then determines a speed change rate to compensate the speed gap therebetween. By referring thereto, the voice speed change part 114 changes the voice speed. Then, the voice recognition part 115 performs voice recognition with respect to the voice-speed-changed voice signal.

[0006] Described next is the operation of the voice recognition device VRAc. The user's sound is captured as command voice together with background noise by the voice input part 111 via a microphone and an amplifier equipped therein, and then an analog signal including the command voice and the background noise is subjected to A/D conversion by an equipped A/D converter. From the voice included in thus obtained digital voice signal, the voice speed calculation part 112 extracts a sound unit which corresponds to the command voice, and calculates the voice speed for the sound unit based on the time taken for the user to produce or utter the sound.

[0007] Here, assuming that the time taken to utter the sound unit (hereinafter, "one-sound unit utterance time") is Ts, and a reference time for utterance of the sound unit (hereinafter, "one-sound unit reference time") is Th. Based thereon, the voice speed change rate determination part 113 determines a speed change rate \( \alpha \) by comparing 1/Ts and 1/Th with each other, which denote a one-sound unit utterance speed and a one-sound unit reference speed, respectively. The speed change rate \( \alpha \) is calculated by the following equation (1).

\[
\alpha = \frac{T_s}{T_h}
\]  

[0008] The equation (1) tells, when the one-sound unit utterance time Ts is shorter than the one-sound unit reference time Th, i.e., when an incoming sound voice speed is faster than that workable by the voice recognition device VRAc, the speed change rate \( \alpha \) is smaller than 1. If this is the case, the incoming command voice should be decreased in speed. Conversely, when the one-sound unit utterance time Ts is longer than the one-sound unit reference time Th, i.e., the incoming command voice speed is slower, the speed change rate \( \alpha \) becomes greater than 1. In such case, the incoming command voice should be increased in speed.

[0009] In the voice recognition device VRAc, the voice speed change part 114 refers to the speed change rate \( \alpha \) to keep the command voice signal constant in speed, and produces a speed-changed command voice signal. The voice recognition part 115 performs voice recognition with respect to the speed-changed command voice signal, and outputs a result obtained thereby.

[0010] Such speed change can be easily done under the recent digital technology. For example, in order to decrease the speed of voice, the voice signal is added with several vowel waveforms having correlation with
the sound unit included in the command voice. To increase the speed of voice, on the other hand, such vowel waveform is decimated from the command voice signal for several times.

[0011] This is a technique for changing the voice speed without affecting the pitch of the command voice. That is, this technique is effective for voice recognition in the case that the user speaks faster or slower than the dictionary voice.

[0012] The above-described conventional voice recognition device VRAc works well for voice recognition when the user’s voice speed is differed to a large extent from the one-sound unit reference speed 1/Th. However, this is not applicable if the user’s voice is differently pitched compared with a reference pitch.

[0013] In detail, although the voice recognition device VRAc can manage with various types of speakers varied in frequency range, i.e., low-pitched man, or high-pitched woman or child. However, voice recognition to be achieved thereby is not satisfactory.

[0014] For the fast speaker speaking in a high speed, it is possible to ask him/her to speak moderately, but is impossible to speak in a different voice pitch. Note that the speaker’s voice pitch is essentially determined by his/her throat especially in shape and size. Since the speaker cannot change his/her throat in shape or size by his/her intention, the voice pitch cannot be changed by his/her intention, too.

[0015] For realizing a voice recognition of various voices with different pitches, the voice recognition device VRAC shall store great number of sample voice data groups each correspond to different speaker such as a man, a woman, or a child speaking in different pitch. Further, the voice recognition device VRAC shall select one group among those great number of sample voice data groups, according to the incoming command voice.

[0016] In order to avoid such nuisance, it seems effective to process the incoming command voice to a pitch optimal for voice recognition. However, since incoming command voices greatly vary in pitches according to the speaker, it is substantially impossible to process the incoming command voice to a desired pitch at one dash. Even in the desired pitch, the correct voice recognition can not be secured, because the content of incoming command voice or a speaking manner may spoil the voice recognition result. As known from this, the pitch considered optical for voice recognition in terms of voice recognition device or sample voice data is not necessarily optimal.

[0017] Therefore, an object of the present invention is to provide a device for normalizing voice pitch to a level considered optimal for voice recognition.

SUMMARY OF THE INVENTION

A first aspect of the present invention is directed to a voice pitch normalization device equipped in a voice recognition device for recognizing an incoming command voice uttered by any speaker based on sample data for a plurality of words, and used to normalize the incoming command voice to be in an optimal pitch for voice recognition, the device comprising:

- a target voice generator for generating a target voice signal by changing said incoming command voice on a predetermined degree basis;
- a probability calculator for calculating a probability indicating a degree of coincidence among the target voice signal and the words in the sample data; and
- a voice pitch changer for repeatedly changing the target voice signal in voice pitch until a maximum of the probabilities reaches a predetermined probability or higher.

[0019] As described above, in the first aspect, an incoming command voice is so adjusted in voice pitch that a probability indicating a degree of coincidence among the incoming command voice and sample voice data for a plurality of words becomes a predetermined value or greater. Therefore, the incoming command voice can be normalized in a fast and correct manner.

[0020] According to a second aspect, in the first aspect, when the maximum of the probabilities is smaller than the predetermined probability, the voice pitch changer includes a voice pitch adjustment for increasing or decreasing the target voice signal on the predetermined degree basis.

[0021] As described above, in the second aspect, the incoming command voice can be normalized even if being lower or higher in voice pitch compared with the sample voice data.

[0022] According to a third aspect, in the second aspect, the voice pitch normalization device further comprises:

- a memory for temporarily storing the incoming command voice;
- a read-out controller for reading out a string of the incoming command voice from the memory, and generating the target voice signal; and
- a read-out clock controller for generating a read-out clock signal with a timing clock determined by frequency, and outputting the timing clock to the memory to change, with the timing specified thereby, the target voice signal in frequency on the predetermined degree basis.

[0023] According to a fourth aspect, in the second aspect, the target voice signal is increased in voice pitch on the predetermined degree basis started from a pitch level of the incoming command voice.

[0024] According to a fifth aspect, in the fourth aspect, the target voice signal is limited in voice pitch up to a first predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the
first predetermined pitch, the target voice signal is decreased in voice pitch on the predetermined degree basis started from the pitch level of the incoming command voice.

As described above, in the fifth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a sixth aspect, in the fifth aspect, the target voice signal is limited in voice pitch down to a second predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the second predetermined pitch, the incoming command voice is stopped being normalized.

As described above, in the sixth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a seventh aspect, in the second aspect, the target voice signal is decreased in voice pitch on the predetermined degree basis started from a pitch level of the incoming command voice.

According to an eighth aspect, in the seventh aspect, the target voice signal is limited in voice pitch down to a third predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the third predetermined pitch, the target voice signal is increased in voice pitch on the predetermined degree basis started from the pitch level of the incoming command voice.

As described above, in the eighth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a ninth aspect, in the eighth aspect, the target voice signal is decreased in voice pitch on the predetermined degree basis started from a second predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the second predetermined pitch, the incoming command voice is stopped being normalized.

As described above, in the ninth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a tenth aspect, in the ninth aspect, the target voice signal is decreased in voice pitch on the predetermined degree basis started from a first predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the first predetermined pitch, the incoming command voice is stopped being normalized.

As described above, in the tenth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to an eleventh aspect, in the tenth aspect, when the maximum of the probabilities is smaller than the predetermined probability, the target voice generator includes a voice pitch adjustment for increasing or decreasing the target voice signal on the predetermined degree basis.

As described above, in the eleventh aspect, the incoming command voice can be normalized even if being lower or higher in voice pitch compared with the sample voice data.

As described above, in the eleventh aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to an twelfth aspect, in the eleventh aspect, the voice recognition device further comprises:

- a memory for temporarily storing the incoming command voice;
- a read-out controller for reading out a string of the incoming command voice from the memory, and generating the target voice signal; and
- a read-out clock controller for generating a read-out clock signal with a timing clock determined by frequency, and outputting the timing clock to the memory to change, with the timing specified thereby, the target voice signal in frequency on the predetermined degree basis.

According to a thirteenth aspect, in the eleventh aspect, the target voice signal is increased in voice pitch on the predetermined degree basis started from a pitch level of the incoming command voice.

As described above, in the thirteenth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a fourteenth aspect, in the thirteenth aspect, the target voice signal is limited in voice pitch up to a first predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the first predetermined pitch, the target voice signal is decreased in voice pitch on the predetermined degree basis started from the pitch level of the incoming command voice.

As described above, in the fourteenth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a fifteenth aspect, in the fourteenth aspect, the target voice signal is limited in voice pitch down to a second predetermined pitch, and when

A tenth aspect of the present invention is directed to a voice recognition device for recognizing an incoming command voice optimally normalized for voice recognition based on sample data for a plurality of words, the device comprising:

- a target voice generator for generating a target voice signal by changing the incoming command voice on a predetermined degree basis;
- a probability calculator for calculating a probability indicating a degree of coincidence among the target voice signal and the words in the sample data; and
- a voice pitch changer for repeatedly changing the target voice signal in voice pitch until a maximum of the probabilities reaches a predetermined probability or higher.

As described above, in the tenth aspect, an incoming command voice is so adjusted in voice pitch that a probability indicating a degree of coincidence among the incoming command voice and sample voice data for a plurality of words becomes a predetermined value or greater. Therefore, the incoming command voice can be normalized in a fast and correct manner.

According to an eleventh aspect, in the tenth aspect, when the maximum of the probabilities is smaller than the predetermined probability, the target voice generator includes a voice pitch adjustment for increasing or decreasing the target voice signal on the predetermined degree basis.

As described above, in the eleventh aspect, the incoming command voice can be normalized even if being lower or higher in voice pitch compared with the sample voice data.

According to a twelfth aspect, in the eleventh aspect, the voice recognition device further comprises:

- a memory for temporarily storing the incoming command voice;
- a read-out controller for reading out a string of the incoming command voice from the memory, and generating the target voice signal; and
- a read-out clock controller for generating a read-out clock signal with a timing clock determined by frequency, and outputting the timing clock to the memory to change, with the timing specified thereby, the target voice signal in frequency on the predetermined degree basis.

According to a thirteenth aspect, in the eleventh aspect, the target voice signal is increased in voice pitch on the predetermined degree basis started from a pitch level of the incoming command voice.

As described above, in the thirteenth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a fourteenth aspect, in the thirteenth aspect, the target voice signal is limited in voice pitch up to a first predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the first predetermined pitch, the target voice signal is decreased in voice pitch on the predetermined degree basis started from the pitch level of the incoming command voice.

As described above, in the fourteenth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

According to a fifteenth aspect, in the fourteenth aspect, the target voice signal is limited in voice pitch down to a second predetermined pitch, and when
the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the second predetermined pitch, the incoming command voice is stopped being normalized.

0042 According to a sixteenth aspect, in the eleventh aspect, the target voice signal is decreased in voice pitch on the predetermined degree basis started from a pitch level of the incoming command voice.

0043 According to a seventeenth aspect, in the sixteenth aspect, the target voice signal is limited in voice pitch down to a third predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the third predetermined pitch, the target voice signal is increased in voice pitch on the predetermined degree basis started from the pitch level of the incoming command voice.

0044 As described above, in the seventeenth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

0045 According to an eighteenth aspect, in the seventeenth aspect, the target voice signal is limited in voice pitch down to a fourth predetermined pitch, and when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the fourth predetermined pitch, the incoming command voice is stopped being normalized.

0046 A nineteenth aspect of the present invention is directed to a voice pitch normalization method utilized for a voice recognition device for recognizing an incoming command voice uttered by any speaker based on sample data for a plurality of words, and applied to normalize the incoming command voice to be in an optimal pitch for voice recognition, the method comprising:

a step of generating a target voice signal by changing the incoming command voice on a predetermined degree basis;
a step of calculating a probability indicating a degree of coincidence among the target voice signal and the words in the sample data; and
a step of repeatedly changing the target voice signal in voice pitch until a maximum of the probabilities reaches a predetermined probability or higher.

0047 As described above, in the nineteenth aspect, an incoming command voice is so adjusted in voice pitch that a probability indicating a degree of coincidence among the incoming command voice and sample voice data for a plurality of words becomes a predetermined value or greater. Therefore, the incoming command voice can be normalized in a fast and correct manner.

0048 According to a twentieth aspect, in the nineteenth aspect, the voice pitch normalization method further comprises a step of, when the maximum of the probabilities is smaller than the predetermined probability, increasing or decreasing the target voice signal on the predetermined degree basis.

0049 As described above, in the twentieth aspect, the incoming command voice can be normalized even if being lower or higher in voice pitch compared with the sample voice data.

0050 According to a twenty-first aspect, in the twentieth aspect, the voice pitch normalization method further comprises:

a step of temporarily storing the incoming command voice;
a step of generating the target voice signal from a string of the temporarily stored incoming command voice; and
a step of determining a timing clock by frequency, in such manner as to change, with the timing specified thereby, the target voice signal in frequency on the predetermined degree basis.

0051 According to a twenty-second aspect, in the twentieth aspect, the voice pitch normalization method further comprises a step of increasing the target voice signal in voice pitch on the predetermined degree basis started from a pitch level of the incoming command voice.

0052 According to a twenty-third aspect, in the twenty-second aspect, the target voice signal is limited in voice pitch up to a first predetermined pitch, and the method further comprises a step of, when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the first predetermined pitch, decreasing the target voice signal in voice pitch on the predetermined degree basis started from the pitch level of the incoming command voice.

0053 As described above, in the twenty-third aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

0054 According to a twenty-fourth aspect, in the twenty-third aspect, the target voice signal is limited in voice pitch down to a second predetermined pitch, and the method further comprises a step of, when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the second predetermined pitch, stopping normalizing the incoming command voice.

0055 As described above, in the twenty-fourth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

0056 According to a twenty-fifth aspect, in the twentieth aspect, the voice pitch normalization method further comprises a step of decreasing the target voice signal in voice pitch on the predetermined degree basis started from a pitch level of the incoming command voice.

0057 According to a twenty-sixth aspect, in the tenth aspect, the voice pitch normalization method further comprises:

a step of temporarily storing the incoming command voice;
a step of generating a target voice signal from a string of the temporarily stored incoming command voice; and
a step of determining a timing clock by frequency, in such manner as to change, with the timing specified thereby, the target voice signal in frequency on the predetermined degree basis.
ty-fifth aspect, the target voice signal is limited in voice pitch down to a third predetermined pitch, and the method further comprises a step of, when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the third predetermined pitch, increasing the target voice signal in voice pitch on the predetermined degree basis started from the pitch level of the incoming command voice.

[0058] As described above, in the twenty-sixth aspect, the capability of the voice recognition device appropriately determines a range for normalizing the incoming command voice.

[0059] According to a twenty-seventh aspect, in the twenty-sixth aspect, the target voice signal is limited in voice pitch down to a fourth predetermined pitch, and the method further comprises a step of, when the maximum of the probabilities failed to reach the predetermined probability or higher before the target voice signal reaching the fourth predetermined pitch, stopping normalizing the incoming command voice.

[0060] These and other objects, features, aspects and advantages of the present invention will become more apparent from the following detailed description of the present invention when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0061] FIG. 1 is a block diagram showing the structure of a voice recognition device equipped with a voice pitch normalization device according to an embodiment of the present invention;
FIG. 2 is a block diagram showing a voice analyzer of FIG. 1 in detail;
FIG. 3 is a diagram showing frequency spectra of voices varied in pitch;
FIG. 4 is a diagram for assistance of explaining exemplary pitch change of voice waveforms, and a pitch change method applied thereto;
FIG. 5 is a flowchart showing the operation of the voice pitch normalization device of FIG. 1;
FIG. 6 is a flowchart showing the detailed operation of the voice pitch normalization device in a maximum probability Pmax (Ni) subroutine shown in FIG. 5; and
FIG. 7 is a block diagram showing the structure of a conventional voice recognition device.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0062] With reference to FIG. 1, described is a voice recognition device incorporated with a device for normalizing voice pitch according to an embodiment of the present invention. A voice recognition device VRAp includes an A/D converter 1, voice pitch normalization device Tr, sample voice data storage 13, and voice analyzer 15. The sample voice data storage 13 stores frequency patterns Psf for each of a plurality of words to be referred to at voice recognition. The frequency patterns Psf are outputted at a predetermined timing. Here, a sound, or voice produced by a user is taken into a voice input means (not shown) composed of a microphone and an amplifier, and is then supplied to the voice recognition device VRAp as an analog signal Sva.

[0063] Such structured voice recognition device VRAp outputs, to a controller 17, a signal Ss which indicates the operating status of the constituents therein. In response thereto, the controller 17 produces a control signal Sc for controlling the operation of those constituents, that is, the comprehensive operation of the voice recognition device VRAp. Note herein that, the operating status signal Ss, the control signal Sc, and the controller 17 are well known, therefore are not described unless otherwise required.

[0064] The A/D converter 1 applies the analog voice signal Sva to A/D conversion, and produces a digital voice signal Svd. The voice pitch normalization device Tr changes the pitch of the digital voice signal Svd by a predetermined level, and produces a pitch-normalized digital voice signal Svc whose pitch is normalized toward an optimal pitch for the voice recognition device VRAp. This pitch-normalized digital voice signal Svc is subject to the voice recognition process to perceive the command the user tried to express his/her intention therefrom. From this point of view, the pitch-normalized digital voice signal Svc is a command voice expressed by a word(s) orally.

[0065] The voice analyzer 15 applies FFT (fast Fourier Transform) to the pitch-normalized digital voice signal Svc, and obtains a frequency pattern Psvc (not shown) thereof. From the sample voice data storage 13, the voice analyzer 15 successively reads out all the sample voice data. Here, the sample voice data is composed of plural pairs of frequency pattern Psf and code Sr, corresponds to different words. The voice analyzer 15 also reads out, from the sample voice data storage 13, the sample voice data for each word. Here, the sample voice data is composed of the frequency pattern Psf and a code Sr. The voice analyzer 15 then compares, for each word, the frequency pattern Psf in the sample voice data with the frequency pattern Psvc of the pitch-normalized digital voice signal Svc. In this manner, calculated is a probability P indicating the degree of coincidence between the frequency pattern Psf and the frequency pattern Psvc.

[0066] The calculation of probability P is made under a conventional technology typified by Hidden Markov Model, which will be described later. Among the probabilities P calculated for all the words found in the sample voice data, the maximum value is referred to as maximum probability Pmax. The code Sr corresponding to the maximum probability Pmax is referred to as maxi-
mum probability code Srp.

[0067] Based on the maximum probability Pmax, the voice pitch normalization device Tr authorizes a word whose frequency pattern Psf coincides with the frequency pattern Psvc as being recognized. For the authorization, a predetermined threshold value called coincidence reference Pth is referred to. Specifically, the voice pitch normalization device Tr determines a word having the maximum probability Pmax greater than the coincidence reference Pth as being coinciding with the incoming command voice. Then, the voice pitch normalization device Tr authorizes that the incoming command voice is recognized correctly.

[0068] When the authorization is established, the voice pitch normalization device Tr outputs a coincidence authorization signal Sj to the voice analyzer 15. In response to the signal Sj, the voice analyzer 15 outputs a maximum probability code Srp indicative of the authorized word (voice authorized sample data). In this sense, the maximum probability code Srp is referred to as recognition code Srp.

[0069] On the other hand, when the maximum probability Pmax is smaller than the coincidence reference Pth in value, the voice pitch normalization device Tr adjusts the digital voice signal Svd in pitch only by a predetermined degree, and thus again produces the pitch-normalized digital voice signal Svc. Then, based thereon, the above procedure is repeated until any word being authorized. Specifically, the comparison in frequency pattern is done for each word in the sample voice data. However, the authorization process is applied only to the word having the maximum probability Pmax.

[0070] Herein, as shown in FIG. 1, the voice pitch normalization device Tr includes memory 3, read-out controller 5, voice pitch optimizer 9, and read-out clock controller 11. The voice pitch optimizer 9 authorizes coincidence between the pitch-normalized digital voice signal Svc and a specific word in the sample voice data on the basis of the maximum probability Pmax provided from the voice analyzer 15.

[0071] To be specific, when the coincidence reference Pth is greater than the maximum probability Pmax in value, the voice pitch optimizer 9 does not authorize such coincidence. If this is the case, the voice pitch optimizer 9 outputs a voice pitch adjusting signal Si to the read-out clock controller 11. This is done to adjust the pitch-normalized digital voice signal Svc provided to the voice pitch normalization device Tr (the voice analyzer 15) by a pitch adjustment degree Ni for the next authorization process.

[0072] Herein, a character / is found in both the pitch adjustment degree Ni and the voice pitch adjusting signal Si is an index specifying the degree for voice pitch adjustment. In this embodiment, although the pitch adjustment index i is exemplarily a positive or a negative integer, it is not restrictive and arbitrary. In this embodiment, the pitch adjustment index i presumably matches, in value, to a pitch adjustment cycle of the pitch-normalized digital voice signal Svc. Herein, the pitch adjustment index i thus denotes the pitch adjustment cycle if necessary.

[0073] In response to the voice pitch adjusting signal Si, the read-out clock controller 11 outputs a read-out clock Scc to the memory 3. This read-out clock Scc changes the pitch-normalized digital voice signal Svc in voice pitch (high or low) by the predetermined degree of Ni.

[0074] The read-out controller 5 monitors the digital voice signal Svd in the memory 3, and produces a read-out control signal Src. The read-out control signal Src so controls the memory 3 as to extract a portion out of the digital voice signal Svd with a timing specified by the read-out clock Scc. The portion is an independent sound unit (s) structuring the incoming command voice included in the digital voice signal Svd, and is read out as the pitch-normalized digital voice signal Svc.

[0075] The memory 3 thus reads out the digital voice signal Svd stored therein with the timing specified by the read-out clock Scc so that the pitch-normalized digital voice signal Svc corresponding to the incoming command is outputted. The pitch-normalized digital voice signal Svc is a signal obtained by changing the digital voice signal Svd in voice pitch by the pitch adjustment degree Ni, which is specified by the voice pitch adjusting signal Si.

[0076] The pitch adjustment degree Ni does not have to be constant but arbitrarily variable. Surely, the capability of the voice recognition device VRAp (especially the combination of the voice analyzer 15 and the sample voice data) naturally determines the acceptable range of the pitch adjustment degree Ni. Hereinafter, such pitch-normalized digital signal Svc adjusted by the pitch adjustment degree Ni is referred to as pitch-normalized digital voice signal Svc(Ni). If required, other signals are also referred to in the same manner.

[0077] With respect to the pitch-normalized digital voice signal Svc whose pitch has been adjusted, the voice analyzer 15 calculates the probability P for every word (M words) in the sample voice data stored in the sample voice data storage 13. Here, M is an arbitrary integer equal to or greater than 1, and equal to the number of codes Sr having the frequency patterns Psf. In this sense, M is the total number of words in sample voice data.

[0078] As shown in FIG. 2, the voice analyzer 15 includes a maximum probability determinator 15a and a coincidence authorized code output 15b. The sample voice data storage 13 outputs a frequency pattern Psf(m) to the maximum probability determinator 15a, and a code Sr(m) corresponding thereto simultaneously to the coincidence authorized code output 15b.

[0079] The coincidence authorized code output 15b retains the value of the code Sr(m) until the next code Sr(m + 1) comes. Herein, m is an arbitrary integer from 1 to M inclusive, and is a parameter indicating any one code or any one of frequency patterns Psf1 to PsfM co-
responding to the M words in the sample voice data stored in the sample voice data storage 13.

[0080] Based on the frequency patterns \( \text{Psf(m)} \) provided by the sample voice data storage 13 and the pitch-normalized digital voice signal \( \text{Svc(Ni)} \), the maximum probability determinator 15a finds a maximum probability \( \text{Pmax(Ni)} \) for that pitch-normalized digital voice signal \( \text{Svc(Ni)} \). Then the maximum probability determinator 15 outputs the maximum probability \( \text{Pmax(Ni)} \) to the voice pitch optimizer 9, and a code retaining signal \( \text{Csr} \) to the coincidence authorized code output 15b.

[0081] In response to the code retaining signal \( \text{Csr} \), the coincidence authorized code output 15b retains the current code \( \text{Sr(m)} \) as an authorization potential code \( \text{Sr'} \). As will be later described, under the condition of the probability \( P \) (i.e., maximum probability \( \text{Pmax(Ni)} \)) being the coincidence reference \( \text{Pth} \) or greater, the code \( \text{Sr} \) for a word having the maximum probability \( \text{Pmax(Ni)} \) is authorized as being the code \( \text{Sr} \) corresponding to the digital voice signal \( \text{Svd} \) equivalent to the incoming command voice (analog voice signal \( \text{Sva} \)). This is the reason why the code \( \text{Sr(m)} \) indicating the maximum probability \( \text{Pmax(Ni)} \) is identified as being the authorization potential code \( \text{Sr'} \). Herein, such authorized code is identified as being the coincidence authorized code \( \text{Sr} \).

[0082] The coincidence authorized code output 15b outputs the coincidence authorized code \( \text{Sr} \) external to the voice recognition device VRAs based on the code retaining signal \( \text{Csr} \) from the maximum probability determinator 15a, the code \( \text{Sr(m)} \) from the sample voice data storage 13, and the coincidence authorization signal \( \text{Sj} \) from the voice pitch optimizer 9. More specifically, after receiving the pitch-normalized digital voice signal \( \text{Svc(Ni)} \), the maximum probability determinator 15a keeps the signal until another pitch-normalized digital voice signal \( \text{Svc(Ni)} \) comes, having adjusted in pitch to a further degree.

[0083] The sample voice data storage 13 successively outputs the previously stored frequency patterns \( \text{Psf} \) (m) corresponding to the words. With every output thereof, the frequency pattern \( \text{Psvc} \) (Ni) of the digital voice signal \( \text{Svc(Ni)} \) is compared to calculate the probability \( P(m) \). If thus calculated probability \( P(m) \) exceeds the probability \( P(m - \beta) \) so far considered maximum, the calculated probability \( P(Ni) \) is updated with the probability \( P(m) \). Here, \( \beta \) is an arbitrary integer from 1 to \( m \) inclusive.

[0084] In response to such update, the maximum probability determinator 15a outputs the code retaining signal \( \text{Csr} \) to the coincidence authorized code output 15b. The code retaining signal \( \text{Csr} \) indicates that the probability \( P(m) \) of the current frequency pattern \( \text{Psf(m)} \) is so far considered maximum. This processing is carried out with respect to all of the frequency patterns \( \text{Psf1} \) to \( \text{PsfM} \) for \( M \) words stored in the sample voice data storage 13, and then the maximum probability \( \text{Pmax(Ni)} \) is determined. Thereafter, the maximum probability \( \text{Pmax(Ni)} \) is outputted to the voice pitch optimizer 9 for authorization. Also, the authorization signal \( \text{Sr} \) (m) for the word indicating the maximum probability \( \text{Pmax(Ni)} \) is stored in the coincidence authorized code output 15b as the authorization potential code \( \text{Sr'} \).

[0085] In the case that the code retaining signal \( \text{Csr} \) is provided by the maximum probability determinator 15a, the current code \( \text{Sr(m)} \) so far considered having the maximum probability \( P \) is retained as the authorization potential code \( \text{Sr'} \) until the next code retaining signal \( \text{Csr} \) comes. When it came, the code \( \text{Sr}(m + \gamma) \) at that time is regarded as the authorization potential code \( \text{Sr'} \). This makes possible for the code \( \text{Sr} \) considered potential as having the maximum probability \( \text{Pmax(Ni)} \) to be always stored as the authorization potential code \( \text{Sr'} \). Herein, \( \gamma \) is an arbitrary integer from 1 to \( (M - m) \) inclusive.

[0086] When the pitch-normalized digital voice signal \( \text{Svc(Ni)} \) is thoroughly compared with every sample voice data (frequency pattern \( \text{Psf(m)} \)) corresponding thereto, the probability \( P \) maximum in value among those found in the maximum probability determinator 15a is outputted to the voice pitch optimizer 9 as the maximum probability \( \text{Pmax(Ni)} \). In the voice pitch optimizer 9, the maximum probability \( \text{Pmax(Ni)} \) is compared with the coincidence reference \( \text{Pth} \).

[0087] When the maximum probability \( \text{Pmax(Ni)} \) is equal to or greater than the coincidence reference \( \text{Pth} \), the voice pitch optimizer 9 outputs the coincidence authorization signal \( \text{Sj} \) to the coincidence authorized code output 15b. The coincidence authorization signal \( \text{Sj} \) authorizes the authorization potential code \( \text{Sr'} \) stored in the coincidence authorized code output 15b as being the coincidence authorized code \( \text{Sr} \). In response thereof, the coincidence authorized code output 15b authorizes the word having the maximum probability \( \text{Pmax(Ni)} \) as having the maximum probability \( \text{Pmax(Ni)} \) with respect to the pitch-normalized digital voice signal \( \text{Svc(Ni)} \) greater than the coincidence reference \( \text{Pth} \).

[0088] In other words, the coincidence authorized code output 15b never outputs the coincidence authorized code \( \text{Sr} \) without receiving the coincidence authorization signal \( \text{Sj} \) from the voice pitch optimizer 9. The coincidence authorized code \( \text{Sr} \) means that the probability \( P \) (maximum probability \( \text{Pmax} \)) with respect to the pitch-normalized digital voice signal \( \text{Svc(Ni)} \) is greater than the coincidence reference \( \text{Pth} \).

[0089] In detail, the voice pitch optimizer 9 compares, with the coincidence reference \( \text{Pth} \), the maximum probability \( \text{Pmax} \) of the code \( \text{Sr} \) corresponding to the pitch-normalized digital voice signal \( \text{Svc(Ni)} \) at the current processing time (i). Then, the voice pitch optimizer 9 determines whether the word (authorization potential code \( \text{Sr'} \)) having the maximum probability \( \text{Pmax} \) at the current processing time (i) has been so far correctly recognized or not. In this case, the authorization potential code \( \text{Sr}(i) \) at the current processing time does not always fall on the authorization potential code \( \text{Sr'}(i-1) \) at the previous processing time.
When the maximum probability $P_{\text{max}}$ is equal to or greater than the coincidence reference $P_{\text{th}}$, the voice pitch optimizer 9 authorizes the coincidence potential code $S_{\text{rp}}$. Therefore, if the voice pitch optimizer 9 does not authorize the coincidence potential code $S_{\text{rp}}$, the coincidence reference $P_{\text{max}}$ among the $M$ words stored in the sample voice data storage 13 is not satisfied for that information. With the coincidence authorization signal $S_{\text{ja}}$ received, the voice analyzer 15 outputs the coincidence potential code $S_{\text{rp}}$ stored therein as the coincidence authorized code $S_{\text{rp}}$.

Next, with reference to FIGS. 3 and 4, described is the basic operational principle of the voice recognition device VRAp.

FIG. 3 shows exemplary frequency spectra (frequency patterns $P_{\text{svc}}$) obtained by subjecting the pitch-normalized digital voice signal $S_{\text{vc}}$ to fast Fourier transform in the voice analyzer 15. In the drawing, a lateral axis indicates frequency $f$, and longitudinal strength $A$. Therein, exemplarily, a one-dot line $L_1$ indicates a typical frequency spectrum of the digital voice signal $S_{\text{vd}}$ including a voice uttered by a man, while a broken line $L_2$ indicates a typical frequency spectrum of the digital voice signal $S_{\text{vd}}$ including a voice uttered by a woman or a child.

A solid line $L_s$ indicates an exemplary frequency spectrum (frequency pattern $P_{\text{sf}}$) of a word (code $S_{r}$) stored in the sample voice data storage 13 as the sample voice data for voice recognition. The word is the one corresponding to the frequency spectra of voices indicated by the lines $L_1$ and $L_2$. Generally, even if the same voice (word) is uttered, as indicated by the one-dot line $L_1$, the frequency spectrum for the man covers the lower frequency region compared with the sample voice. On the other hand, as indicated by the broken line $L_2$, that for the woman or child covers the higher frequency region.

By taking such frequency spectra into consideration, the voice analyzer 15 goes through comparison between the frequency pattern $P_{\text{svc}}$ of the pitch-normalized digital voice signal $S_{\text{vc}}$ typified by the line $L_1$ or $L_2$, and the frequency patterns $P_{\text{sf}}(m)$ for every word ($S_{r}(m)$) in the sample voice data typified by the line $L_s$. Then, the degree of coincidence $P(m)$ is calculated for every word ($S_{r}(m)$). Such calculation of the probability $P(m)$ can be done under the conventional technology such as Hidden Markov Model.

The sample voice data ($L_s$) stored in the sample voice data storage 13 is often set to be in the middle level of the man's voice ($L_1$) and the woman's voice ($L_2$). Therefore, if their voices are extremely high or low, the frequencies ($L_1$, $L_2$) thereof are different from that of the sample voice data ($L_s$) to a greater degree. Consequently, even if the word is correct, the probability $P$ thereof cannot reach the coincidence reference $P_{\text{th}}$, rendering voice correction failed.

Therefore, in the present invention, if the maximum probability $P_{\text{max}}(m)$ among the $M$ words stored in the sample voice data does not satisfy the coincidence reference $P_{\text{th}}$, the pitch level of the pitch-normalized digital voice signal $S_{\text{vc}}$ is regarded as the reason. Thus, the pitch level is adjusted (high or low).

To be specific, when the maximum probability $P_{\text{max}}(m)$ detected by the voice analyzer 15 is determined as being not reaching the coincidence reference $P_{\text{th}}$ by the voice pitch optimizer 9, the voice pitch adjusting signal $S_{\text{is}}$ is outputted to the read-out clock controller 11. The voice pitch adjusting signal $S_{\text{is}}$ has been so set as to adjust the pitch-normalized digital voice signal $S_{\text{vc}}$ in voice pitch by the predetermined degree of $N_\text{i}$.

As described in the foregoing, the memory 3 outputs the pitch-normalized digital voice signal $S_{\text{vc}}(N_\text{i})$ which has been adjusted in voice pitch by the degree of $N_\text{i}$ to the voice analyzer 15. Therein, the pitch-normalized digital voice signal $S_{\text{vc}}(N_\text{i})$ is subjected to the above-described voice analysis so that the maximum probability $P_{\text{max}}$ is calculated. In this case, the word which has indicated the maximum probability $P_{\text{max}}(i-1)$ during voice analysis at the previous processing time ($i-1$) does not necessarily indicate the maximum probability $P_{\text{max}}(i)$ at the current processing time ($i$).

This is because, as described by referring to FIG. 3, the probability $P(m)$ varies to a considerable degree depending on the proximity between the frequency pattern $P_{\text{svc}}(N_\text{i})$ of the pitch-normalized digital voice signal $S_{\text{vc}}(N_\text{i})$ exemplarily indicated by the lines $L_1$ or $L_2$, and the frequency pattern $P_{\text{sf}}(m)$ of the sample voice exemplarily indicated by the line $L_s$. As a result, when the proximity of voice pitch is insufficient, the word not corresponding to the pitch-normalized digital voice signal $S_{\text{vc}}$ may become erroneously higher in probability $P$ compared with the word corresponding thereto.

Here, the closer the proximity of voice pitch, the greater the possibility $P$ of the correct word. Focusing on this respect in this invention, the coincidence reference $P_{\text{th}}$ is set according to the capability of the voice recognition device VRAp. When the maximum probability $P_{\text{max}}$ is equal to or greater than the coincidence reference $P_{\text{th}}$, the word corresponding thereto is authorized as being correctly recognized by voice.

That is, in the present invention, the pitch of the pitch-normalized digital voice signal $S_{\text{vc}}$ is normalized through adjustment until the maximum probability $P_{\text{max}}$ satisfies the coincidence reference $P_{\text{th}}$. In this manner, finding the correct word is based not on every word but only on the maximum probability $P_{\text{max}}$, whereby load on data processing is considerably lessened. Also, every single word included in the sample voice data is targeted for voice recognition, thereby rendering voice recognition fast and correct.

With reference to FIG. 4, the voice pitch normalization device $T_r$ (read-out clock controller 11) is described for its pitch change to a further degree. In the drawing, a lateral axis indicates time $t$, and longitudinal voice strength $A$. A waveform $W_S$ shows the change of a voice waveform (frequency $P_{\text{sf}}(m)$) with time stored in the sample voice data storage 13.
procedure then goes to step S6.

[0109] In step S6, the A/D converter 1 subjects the analog voice signal Sva to A/D conversion. Then, thus produced digital voice signal Svd is outputted to the memory 3. The procedure goes to step S8.

[0110] In step S8, the memory 3 stores every incoming digital voice signal Svd. The procedure then goes to step S10.

[0111] In step S10, the read-out controller 5 monitors the memory 3 for its input status to judge whether the speaker's voice input (analog voice signal Sva) has been through. In this judgement, for example, a length of time having no input of analog voice signal Sva is referred to to see whether reaching a predetermined reference threshold value. Alternatively, the speaker may use some appropriate means to inform the voice recognition device VRAp or the voice pitch normalization device Tr that the signal input is now through.

[0112] If the speaker keeps speaking, the judgement is No, therefore the procedure returns to step S4 to repeat steps S4, S6, and S8 for inputting the speaker's voice, generating the digital voice signal Svd, and storing the signal in the memory 3. Once the analog voice signal Sva which is an independent voice string structured by one or more sound units uttered by the speaker was completely inputted, the determination becomes Yes. This means that the memory 3 is through with storing the digital voice signal Svd including the voice uttered by the speaker. Therefore, the procedure goes to step S12.

[0113] In step S12, the read-out controller 5 refers to the memory 3 for the digital voice signal Svd and the read-out clock Scc stored therein to read out the pitch-normalized digital voice signal Svc(Ni). Here, the pitch-normalized digital voice signal Svc(Ni) is obtained by adjusting (increasing or decreasing) the digital voice signal Svd in voice pitch by a predetermined degree by Ni, which is equivalent to the voice pitch adjusting signal Si referred to for generating the read-out clock Scc.

[0114] Note herein that, if the pitch-normalized digital voice signal Svc(Ni) is read out from the memory 3 for the first time, the pitch adjustment degree is 0 as the index i has been initialized in step S2. In other words, the digital voice signal Svd is read out as the pitch-normalized digital voice signal Svc(Ni) without being adjusted by voice pitch. The procedure then goes to step S14.

[0115] In step S14, as to the pitch-normalized digital voice signal Svc(Ni) thus adjusted in voice pitch by the degree Ni specified by the index i, the voice analyzer 15 subjects the signal to Fourier transform so that a frequency pattern Psvc(Ni) is produced. Thereafter, the frequency spectrum analysis is carried out. The procedure then goes to step #100 for maximum probability Pmax(Ni) detection subroutine.

[0116] In step #100, the frequency pattern Psvc(Ni) of the pitch-normalized digital voice signal Svc(Ni) is compared with the frequency pattern Psf(m) being the sample voice data for each word read out from the sample voice data storage 13, and then the probability P(m) in-
indicating the degree of coincidence therebetween is detected. Such technique for comparing the patterns of
digital voice signal and sample voice data with each other for calculating the probability P is typified by the Hidden
Markov Model, which is the known technique.

With reference to FIG. 6, described next is the detailed operation in step #100. Once the maximum probability
Pmax(Ni) subroutine in step #100 was started,

first, in step S102, from the memory 3, the frequency pattern Psvc(Ni) of the pitch-normalized digital voice
signal Svc(Ni) is provided to the maximum probability
determinator 15a in the voice analyzer 15. The procedure
then goes to step S104.

In step S104, the voice analyzer 15 is initialized. Specifically, in the maximum probability determinator 15a, m is set to 1 and the maximum probability Pmax(Ni) to 0. Moreover, in the coincidence authorization code output 15b, the authorization potential code Srp' is set to 0. The procedure then goes to step S106.

In step S106, from the sample voice data storage 13, the frequency pattern Psf(m) and code Sr(m) are inputted into the maximum probability determinator 15a and the coincidence authorization code output 15b, respectively. The procedure then goes to step S108.

In step S108, the maximum probability determinator 15a calculates the probability P(m) indicating the degree of coincidence between the frequency pattern Psvc(Ni) inputted in step S102 and the frequency pattern Psf(m) received in step S106. The procedure then goes to step S110.

In step S110, the maximum probability determinator 15a determines whether or not the maximum probability P(m) is equal to or greater than the maximum probability Pmax. If Yes, the procedure goes to step S112.

In step S112, the current probability P(m) in the maximum probability determinator 15a is set to the maximum probability Pmax(Ni). The procedure then goes to step S114.

In step S114, the maximum probability determinator 15a outputs a code retaining signal Csr to the coincidence authorization code output 15b. The procedure then goes to step S116.

In step S116, the coincidence authorization code output 15b sets, responding to the code retaining signal Csr, the code Sr(m) currently stored therein to the authorization potential code Srp'. The procedure then goes to step S118.

On the other hand, if it is determined as No in step S110, that is, if the probability P(m) is determined as being smaller than the maximum probability Pmax, the procedure skips steps S112, S114, and S116 and goes to step S118.

In step S118, determination is made whether m is equal to M. In the case that m is smaller than M, it is determined as No, and then the procedure goes to step S120.

In step S120, m is incremented by 1, and then the procedure returns to step S106. Thereafter, the processing in steps S106 to S120 is repeated until determination made in step S118 becomes Yes by m becoming equal to M through increment.

Determined in step S118 is the probabilities P(m) for the frequency patterns Psf(1) to Psf(M) in the sample voice data stored in the sample voice data storage 13, and which of the calculated probabilities P(m) is the maximum probability Pmax. As such, with respect to every authorization signal Sr stored in the sample voice data storage 13, calculated is the maximum probability Pmax and the authorization potential code Srp'. Then, the procedure goes to step S122.

In step S122, the maximum probability Pmax(15a) outputs the maximum probability Pmax(Ni) internally stored therein in step S112 to the voice pitch optimizer 9.

In this manner, the voice analyzer 15 looks for the probability P highest among those for the sample voice data (voice frequency patterns Psf) and the voice signal (pitch-normalized digital voice signal Svc) including the incoming command voice (analog voice signal Sva), and then outputs only the sample voice data (coincidence authorization code Srp) showing the maximum Pmax(Ni). This is the end of step #100.

In step S18, the voice pitch optimizer 9 determines whether the maximum probability Pmax(Ni) is equal to or greater than the coincidence reference Pth. In the case that the maximum probability Pmax(Ni) is smaller than the coincidence reference Pth, that is, when it is not sufficient to determine whether voice recognition is correctly done even if the sample voice data shows the highest probability P at the current processing time (i), determined is No and the procedure goes to step S20.

In step S20, referred to is a maximum pitch flag FNmax showing whether the pitch adjustment degree Ni for the pitch-normalized digital voice signal Svc(Ni) has reached an allowable maximum pitch Nmax. In the case that the maximum pitch flag FNmax is not 1, that is, when the pitch adjustment degree Ni does not yet reach the maximum pitch flag FNmax, determined is No, and the procedure goes to step S22.

In step S22, determined is whether the pitch adjustment degree Ni is equal to or greater than the allowable maximum pitch Nmax. If determined is No, the procedure goes to step S24.

In step S24, the index i for adjusting the voice pitch is incremented by 1. This means that the pitch adjustment degree Ni is increased (put higher). The procedure then goes to step S26.

In step S26, the voice pitch optimizer 9 produces a voice pitch adjusting signal Si for output to the readout clock controller 11. Thereafter, the procedure returns to step S12.

In step S22, is Yes, that is, when the pitch adjustment degree Ni is determined as having reached the allowable maximum
pitch \( N_{\text{max}} \), the procedure goes to step S28.

[0137] In step S28, the maximum pitch flag \( F_{\text{Nmax}} \) is set to 1. The procedure then goes to step S30.

[0138] In step S30, the index \( i \) for adjusting the voice pitch is reset to 0. The procedure then goes to step S32.

[0139] In step S32, determined is whether the pitch adjustment degree \( N_i \) is equal to or smaller than an allowable minimum pitch \( N_{\text{min}} \). If determined is No, the procedure goes to step S34.

[0140] In step S34, the index \( i \) is decremented by 1. This means that the pitch adjustment degree \( N_i \) is decreased (put lower). To be more specific, compared with the digital voice signal \( S_{\text{vd}} \), the pitch normalized digital voice signal \( S_{\text{vc}}(N_i) \) is decreased in pitch voice to be lower by the pitch adjustment degree \( N_i \). The procedure then goes to step S26.

[0141] On the other hand, if determined in step S32 is Yes, that is, when the pitch adjustment degree \( N_i \) is determined as being the allowable minimum pitch \( N_{\text{min}} \) or smaller, this is the end of the procedure. This indicates that the analog voice signal \( S_{\text{va}} \) has not been voice recognizable.

[0142] In the case that determined in step S20 is Yes, that is, when the maximum pitch flag \( F_{\text{Nmax}} \) is 1 (set in step S28), the procedure goes to step S32.

[0143] In the case that determined in step S18 is Yes, that is, when the maximum probability \( P_{\text{max}}(N_i) \) is equal to or greater than the coincidence reference \( P_{\text{th}} \), this indicates that the word \( (S_{\text{rp}}) \) corresponding thereto is correct. The procedure then goes to step S36.

[0144] In step S36, the maximum probability determinator 15a outputs the coincidence authorization signal \( S_{\text{j}} \) to the coincidence authorization code output 15b. The procedure then goes to step S38.

[0145] In response to the coincidence authorization signal \( S_{\text{j}} \), the coincidence authorization code output 15b outputs, externally to the voice recognition device VRAp, the authorization potential code \( S_{\text{rp}'}, \) set in step S116 (#100) as the coincidence authorization code \( S_{\text{rp}} \). This is the end of the operation of the voice recognition device VRAp.

[0146] By referring to the above-described flowcharts, the operation of the voice recognition device VRAp is described in a specific manner. Once the voice recognition device VRAp is started for its operation of voice recognition, the voice pitch normalization device \( T_r \) is initiated in step S2. Accordingly, the pitch adjustment index \( i \) is set to 0, and the allowable maximum pitch \( N_{\text{max}} \) and the allowable minimum pitch \( N_{\text{min}} \) are each set to a predetermined value.

[0147] In steps S4, S6, S8, and S10, the speaker's voice is stored in the memory 3 as the digital voice signal \( S_{\text{vd}} \).

[0148] In step S12, the digital voice signal \( S_{\text{vd}} \) is read from the memory 3 according to the read-out clock \( S_{\text{cc}} \) (i) which corresponds to the index \( i =0 \) initialized in step S2. Accordingly, the pitch-normalized digital voice signal \( S_{\text{vc}}(N_i) \) is outputted to the voice analyzer 15.

Here, since \( i =0 \), the pitch adjustment degree \( N_i =0 \), and the pitch-normalized digital voice signal \( S_{\text{vc}}(N_i) \) is equal in voice pitch to the digital voice signal \( S_{\text{vd}} \).

[0149] The voice analyzer 15 carries out the frequency spectrum analysis with respect to the pitch-normalized digital voice signal \( S_{\text{vc}}(N_i) \) (S14). Moreover, the probabilities \( P(1) \) to \( P(M) \) are detected among the frequency patterns \( Psf(i) \) of the pitch-normalized digital voice signal \( S_{\text{vc}}(N_i) \) at \( i =0 \) and the frequency patterns \( Psf(1) \) to \( Psf(M) \) of the sample voice data read from the sample voice storage 13. Thereafter, the sample voice data (authorization potential code \( S_{\text{rp}'} \)) showing the highest probability \( P \) thereamong is looked for so that the maximum probability \( P_{\text{max}} \) is calculated. In this manner, the maximum probability \( P_{\text{max}}(N_i) \) corresponding to the current pitch adjustment degree \( N_i \) is produced (#100).

[0150] When the maximum probability \( P_{\text{max}}(N_i) \) is equal to or greater than the coincidence reference \( P_{\text{th}} \), the voice pitch optimizer 9 authorizes the voice data (authorization potential code \( S_{\text{rp}} \)) of the word showing the maximum probability \( P_{\text{max}} \) as coinciding with the digital voice signal \( S_{\text{vd}} \), i.e., the speaker's voice \( (S_{\text{18}}) \). The voice pitch optimizer 9 also outputs the coincidence authorization signal \( S_{\text{j}} \) (S36) so as to bring the voice analyzer 15 to output the authorization potential code \( S_{\text{rp}'} \) as the coincidence authorized code \( S_{\text{rp}} \) (S38).

[0151] On the other hand, when the maximum probability \( P_{\text{max}}(N_i) \) is smaller than the coincidence reference \( P_{\text{th}} \), determined in step S18 is that the voice recognition is not correctly done regardless of the sample voice data showing the highest probability \( P \) at that time. Then, in step S20, determination is made whether the pitch adjustment degree \( N_i \) has reached its upper limit (i.e., whether the pitch has been adjustable increased) with reference to the maximum pitch flag \( F_{\text{Nmax}} \) for read-out of the pitch-normalized digital voice signal \( S_{\text{vc}}(N_i) \) from the digital voice signal \( S_{\text{vd}} \). If determined as not yet, confirmation is made in step S22 that the pitch adjustment degree \( N_i \) has not yet reached the allowable maximum pitch \( N_{\text{max}} \). Then, in step S24, the index \( i \) for adjusting the voice pitch is incremented by 1. On the basis of the voice pitch adjusting signal \( S_{\text{i}} \) indicating the incremented index \( i \), the read-out clock \( S_{\text{cc}} \) is produced for output to the memory 3.

[0152] In step S12, according to the read-out clock \( S_{\text{cc}} \), the memory 3 outputs the pitch-normalized digital voice signal \( S_{\text{vc}}(N_i) \), whose voice pitch is increased by the degree of \( N_i \) specified for the digital voice signal \( S_{\text{vd}} \) by the index \( i \). Thereafter, the processing in steps S20 to S34 is repeated until determination made in step S18 becomes Yes, that is, until the maximum probability \( P_{\text{max}} \) is determined as being equal to or greater than the coincidence reference \( P_{\text{th}} \).

[0153] To be more specific, until the pitch adjustment degree \( N_i \) is determined as having reached the allowable maximum pitch \( N_{\text{max}} \) in step S22, unless determination made in step S18 becomes Yes, the loops each
composed of steps S20 to S26, and S12 to S18 are repeated. In this manner, for every pitch-normalized digital voice signal $S_{vc}(Ni)$ whose voice pitch is increased by the predetermined degree of $Ni$ (S24, S26, S12), the maximum probability $P_{max}$ (S14, #100) is calculated.

[0154] During such processing, for every increase in pitch of the pitch-normalized digital voice signal $S_{vc}(Ni)$ by degree of $Ni$, the sample voice data showing the maximum probability $P_{max}$ may change. In detail, the sample data showing the maximum probability $P_{max}$ at the previous processing time $(i-1)$ does not necessarily show the maximum probability $P_{max}$ at the current processing time $(i)$. As such, for every increase by the predetermined degree of $Ni$, the maximum probability $P_{max}$ of the targeted pitch-normalized digital voice signal $S_{vc}(Ni)$ is compared with the coincidence reference $P_{th}$. If the maximum probability $P_{max}$ is equal to or greater than the coincidence reference $P_{th}$, voice recognition is determined as having done under the best condition, and thus the code $Sr$ corresponding to the sample voice data showing the maximum probability $P_{max}$ is outputted as the coincidence authorized code $Sr_p$.

[0155] As is known from the above, according to the present invention, a condition for optimal voice recognition is set only to the maximum probability $P_{max}$. In this manner, until such condition is satisfied, the pitch adjustment of the pitch-normalized digital voice signal $S_{vc}$ is done by taking all of the sample voice data into consideration regardless of the probability $P$ thereof. In this embodiment, a voice pitch of an incoming analog voice signal $S_{va}$ (digital voice signal $S_{vd}$) is taken as a reference $(i=0)$ so that increase in voice pitch is firstly done (S22, S24, S26) by the predetermined degree of $Ni$. Then, until the condition is determined as being satisfied (S12, S14, #100) (No in step S18), the pitch is increased up to the allowable maximum pitch $N_{max}$ (S22).

[0156] In the case that the condition is not determined as being satisfied (No in S18) even if the pitch is increased up to the allowable maximum pitch $N_{max}$, the pitch adjustment is done in a decreasing adjustment mode this time. The mode can be switched by setting the maximum pitch flag $F_{N_{max}}$ to 1 (S28) and the index $i$ for adjusting the voice pitch to 0 (S30).

[0157] In the decreasing adjustment mode, the maximum pitch flag $F_{N_{max}}$ is 1 (S20), thereby skipping the processing of increasing the voice pitch (S22, S24). Here, until the pitch adjustment degree $Ni$ reaches the allowable minimum pitch $N_{min}$ (No in step S32), the index $i$ is decremented by 1 (S34) so that the voice pitch adjusting signal $Si$ is produced (S34).

[0158] As a result of such processing, decreasing in pitch is firstly done by the predetermined degree of $Ni$ by taking the pitch of the analog voice signal $S_{va}$ (digital voice signal $S_{vd}$) as a reference $(i=0)$ (S32, S34, S26, S12, S14, #100). Then, until the condition for optimal voice recognition is determined as being satisfied (No in step S18), the pitch is decreased down to the allowable minimum pitch $N_{min}$. If the maximum probability $P_{max}$ is not determined as being equal to or greater than the coincidence reference $P_{th}$ (Yes in step S18) in the modes of increasing and decreasing the voice pitch, the processing is terminated Yes in S32.

[0159] In this embodiment, the pitch-normalized digital voice signal $S_{vc}$ is first increased in pitch starting from the pitch level of the digital voice signal $S_{vd}$ up to the allowable maximum pitch $N_{max}$. Note herein that, thereafter, the pitch of the pitch-normalized digital voice signal $S_{vc}$ increased up to the allowable maximum pitch $N_{max}$ is put back to the pitch level of the digital voice signal $S_{vd}$, and then is started to be decreased down to the allowable minimum pitch $N_{min}$. However, decreasing first and then increasing the voice pitch is easier than the above disclosure.

[0160] Alternatively, the pitch-normalized digital voice signal $S_{vc}$ may be increased in pitch first all the way to the allowable maximum pitch $N_{max}$, and then decreased down to the allowable minimum pitch $N_{min}$ by degrees. This is also easier than the above disclosure.

[0161] Or, instead of the range between the allowable minimum pitch $N_{min}$ and the allowable maximum pitch $N_{max}$ applied to the pitch adjustment, applied may be a range between the pitch level of the digital voice signal $S_{vd}$ and the allowable minimum pitch $N_{min}$, or a range between the pitch level of the digital voice signal $S_{vd}$ and the allowable maximum pitch $N_{max}$. This is also easier than the above disclosure.

[0162] As described in the foregoing, in the present invention, the voice pitch is normalized through repeated adjustment under the condition of the maximum probability $P_{max}$ satisfying the coincidence reference $P_{th}$. In this manner, while taking every word in the sample voice data into consideration for voice recognition, the maximum probability $P_{max}$ is only referred to for word selection. Accordingly, data processing is considerably lessened by load, successfully leading to fast and correct voice recognition.

[0163] While the invention has been described in detail, the foregoing description is in all aspects illustrative and not restrictive. It is understood that numerous other modifications and variations can be devised without departing from the scope of the invention.

Claims

1. A voice pitch normalization device (Tr) equipped in a voice recognition device (VRAp) for recognizing an incoming command voice ($S_{va}$) uttered by any speaker based on sample data ($Ps_{f}$) for a plurality of words, and used to normalize the incoming command voice ($S_{vc}$) to be in an optimal pitch for voice recognition, the device comprising:

- target voice generation means (9, 11, 3, 5; S24, S34, S26, S12) for generating a target voice
signal (Svc(Ni)) by changing said incoming command voice (Svd) on a predetermined degree (Ni) basis;
probability calculation means (15; S14, #100) for calculating a probability (P) indicating a degree of coincidence among said target voice signal (Svc(Ni)) and the words in said sample data (Psf); and
voice pitch change means (9, 11, 3, 5; S18, S20, S32) for repeatedly changing said target voice signal (Svc(Ni)) in voice pitch until a maximum (Pmax) of said probabilities (P) reaches a predetermined probability (Pth) or higher.

2. The voice pitch normalization device (Tr) as claimed in claim 1, wherein, when the maximum (Pmax) of said probabilities (P) is smaller than said predetermined probability (Pth), said voice pitch change means (9, 11, 3, 5) includes voice pitch adjustment means (S24, S34) for increasing or decreasing said target voice signal (Svc(Ni)) on the predetermined degree (Ni) basis.

3. The voice pitch normalization device (Tr) as claimed in claim 2, further comprising:

memory means (3; S8) for temporarily storing said incoming command voice (Svd);
read-out control means (5, S4, S6, S8) for reading out a string of said incoming command voice (Svd) from said memory means (3), and generating the target voice signal (Svc); and
read-out clock control means (11, S22) for generating a read-out clock signal (Scc) with a timing clock determined by frequency, and outputting the timing clock (Scc) to said memory means (3) to change, with the timing specified thereby, said target voice signal (Svc(Ni)) in frequency on said predetermined degree (Ni) basis.

4. The voice pitch normalization device (Tr) as claimed in claim 2, wherein said target voice signal (Svc(Ni)) is increased in voice pitch on said predetermined degree (Ni) basis started from a pitch level of said incoming command voice (Svd).

5. The voice pitch normalization device (Tr) as claimed in claim 4, wherein said target voice signal (Svc(Ni)) is limited in voice pitch up to a first predetermined pitch (Mmax), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the first predetermined pitch (Mmax), said target voice signal (Svc(Ni)) is decreased in voice pitch on said predetermined degree (Ni) basis started from the pitch level of said incoming command voice (Svd).

6. The voice pitch normalization device (Tr) as claimed in claim 5, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a second predetermined pitch (Mmin), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the second predetermined pitch (Mmin), said incoming command voice is stopped being normalized.

7. The voice pitch normalization device (Tr) as claimed in claim 2, wherein said target voice signal (Svc(Ni)) is decreased in voice pitch on said predetermined degree (Ni) basis started from a pitch level of said incoming command voice (Svd).

8. The voice pitch normalization device (Tr) as claimed in claim 7, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a third predetermined pitch (Mmin), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the third predetermined pitch (Mmin), said target voice signal (Svc(Ni)) is increased in voice pitch on said predetermined degree (Ni) basis started from the pitch level of said incoming command voice (Svd).

9. The voice pitch normalization device (Tr) as claimed in claim 8, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a fourth predetermined pitch (Mmax), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the fourth predetermined pitch (Mmax), said incoming command voice is stopped being normalized.

10. A voice recognition device (VRAp) for recognizing an incoming command voice (Sva) optimally normalized for voice recognition based on sample data (Psf) for a plurality of words, the device comprising:
target voice generation means (9, 11, 3, 5; S24, S34, S26, S12) for generating a target voice signal (Svc(Ni)) by changing said incoming command voice (Svd) on a predetermined degree (Ni) basis;
probability calculation means (15; S14, #100) for calculating a probability (P) indicating a degree of coincidence among said target voice signal (Svc(Ni)) and the words in said sample data (Psf); and
voice pitch change means (9, 11, 3, 5; S18, S20, S32) for repeatedly changing said target voice signal (Svc(Ni)) in voice pitch until a maximum (Pmax) of said probabilities (P) reaches a predetermined probability (Pth) or higher.
11. The voice recognition device (VRAp) as claimed in claim 10, wherein, when the maximum (Pmax) of said probabilities (P) is smaller than said predetermined probability (Pth), said target voice generation means includes voice pitch adjustment means (S24, S34) for increasing or decreasing said target voice signal (Svc(Ni)) on the predetermined degree (Ni) basis.

12. The voice recognition device (VRAp) as claimed in claim 11, further comprising:

memory means (3; S8) for temporarily storing said incoming command voice (Svd); read-out control means (5, S4, S6, S8) for reading out a string of said incoming command voice (Svd) from said memory means (3), and generating the target voice signal (svd); and read-out clock control means (11, S22) for generating a read-out clock signal (Scc) with a timing clock determined by frequency, and outputting the timing clock (Scc) to said memory means (3) to change, with the timing specified thereby, said target voice signal (Svc(Ni)) in frequency on said predetermined degree (Ni) basis.

13. The voice recognition device (VRAp) as claimed in claim 11, wherein said target voice signal (Svc(Ni)) is increased in voice pitch on said predetermined degree (Ni) basis started from a pitch level of said incoming command voice (Svd).

14. The voice recognition device (VRAp) as claimed in claim 13, wherein said target voice signal (Svc(Ni)) is limited in voice pitch up to a first predetermined pitch (Mmin), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the first predetermined pitch (Mmin), said target voice signal (Svc(Ni)) is decreased in voice pitch on said predetermined degree (Ni) basis started from the pitch level of said incoming command voice (Svd).

15. The voice recognition device (VRAp) as claimed in claim 14, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a second predetermined pitch (Mmin), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the second predetermined pitch (Mmin), said incoming command voice is stopped being normalized.

16. The voice recognition device (VRAp) as claimed in claim 11, wherein said target voice signal (Svc(Ni)) is decreased in voice pitch on said predetermined degree (Ni) basis started from a pitch level of said incoming command voice (Svd).

17. The voice recognition device (VRAp) as claimed in claim 16, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a third predetermined pitch (Mmin), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the third predetermined pitch (Mmin), said target voice signal (Svc(Ni)) is increased in voice pitch on said predetermined degree (Ni) basis started from the pitch level of said incoming command voice (Svd).

18. The voice recognition device (VRAp) as claimed in claim 17, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a fourth predetermined pitch (Mmax), and when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the fourth predetermined pitch (Mmax), said incoming command voice is stopped being normalized.

19. A voice pitch normalization method utilized for a voice recognition device (VRAp) for recognizing an incoming command voice (Sva) uttered by any speaker based on sample data (Psf) for a plurality of words, and applied to normalize the incoming command voice (Sva) to be in an optimal pitch for voice recognition, the method comprising:

a step (S24, S34, S26, S12) of generating a target voice signal (Svc(Ni)) by changing said incoming command voice (Svd) on a predetermined degree (Ni) basis;
a step (S14, #100) of calculating a probability (P) indicating a degree of coincidence among said target voice signal (Svc(Ni)) and the words in said sample data (Psf); and a step (S18, S20, S32) of repeatedly changing said target voice signal (Svc(Ni)) in voice pitch until a maximum (Pmax) of said probabilities (P) reaches a predetermined probability (Pth) or higher.

20. The voice pitch normalization method as claimed in claim 19, further comprising a step (S24, S34) of, when the maximum (Pmax) of said probabilities (P) is smaller than said predetermined probability (Pth), increasing or decreasing said target voice signal (Svc(Ni)) on the predetermined degree (Ni) basis.

21. The voice pitch normalization method as claimed in claim 20, further comprising:

a step (S8) of temporarily storing said incoming
command voice (Svd);
a step (S4, S6, S8) of generating the target voice signal (svd) from a string of said temporarily stored incoming command voice (Svd);
and
a step (S22) of determining a timing clock by frequency, in such manner as to change, with the timing specified thereby, said target voice signal (Svc) in frequency on said predetermined degree (Ni) basis.

22. The voice pitch normalization method as claimed in claim 20, further comprising a step (S24) of increasing said target voice signal (Svc(Ni)) in voice pitch on said predetermined degree (Ni) basis started from a pitch level of said incoming command voice (Svd).

23. The voice pitch normalization method as claimed in claim 22, wherein said target voice signal (Svc(Ni)) is limited in voice pitch up to a first predetermined pitch (Mmax), and
the method further comprises a step (S30, S34) of, when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the first predetermined pitch (Mmax), decreasing said target voice signal (Svc(Ni)) in voice pitch on said predetermined degree (Ni) basis started from the pitch level of said incoming command voice (Svd).

24. The voice pitch normalization method as claimed in claim 23, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a second predetermined pitch (Mmin), and
the method further comprises a step (S18, S20, S32) of, when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the second predetermined pitch (Mmin), stopping normalizing said incoming command voice.

25. The voice pitch normalization method as claimed in claim 20, further comprising a step (S34) of decreasing said target voice signal (Svc(Ni)) in voice pitch on said predetermined degree (Ni) basis started from a pitch level of said incoming command voice (Svd).

26. The voice pitch normalization method as claimed in claim 25, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a third predetermined pitch (Mmin), and
the method further comprises a step (S24) of, when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the third predetermined pitch (Mmin), increasing said target voice signal (Svc(Ni)) in voice pitch on said predetermined degree (Ni) basis started from the pitch level of said incoming command voice (Svd).

27. The voice pitch normalization method as claimed in claim 26, wherein said target voice signal (Svc(Ni)) is limited in voice pitch down to a fourth predetermined pitch (Mmax), and
the method further comprises a step (S32) of, when said maximum (Pmax) of said probabilities (P) failed to reach said predetermined probability (Pth) or higher before the target voice signal (Svc(Ni)) reaching the fourth predetermined pitch (Mmax), stopping normalizing said incoming command voice.

Patentansprüche

1. Stimmlagen-Normalisierungsvorrichtung (Tr) ausgestattet in einer Spracherkennungsvorrichtung (VRAp) zum Erkennen einer eingehenden Kommandostimme (Sva), welche von irgendeinem Sprecher geäußert wird, basierend auf Musterdaten (Psf) für eine Vielzahl von Wörtern, und verwendet, um die eingehende Kommandostimme (Sva) zu normalisieren, um in einer optimalen Lage für Spracherkennung zu sein, wobei die Vorrichtung umfasst:

-Zielprachterzeugungsmittel (9, 11, 3, 5; S24, S34, S26, S12) zum Erzeugen eines Zielsprachsignals (Svc(Ni)) durch Verändern der eingehenden Kommandostimme (Sva) auf Basis eines vorherbestimmt Grades (Ni);
-Wahrscheinlichkeitsberechnungsmittel (15; S14, #100) zum Berechnen einer Wahrscheinlichkeit (P), welche anzegend für einen Grad der Übereinstimmung unter dem Zielsprachsignal (Svc(Ni)) und den Wörtern in den Musterdaten (Psf) ist; und
-Stimmlagenveränderungsmittel (9, 11, 3, 5; S18, S20, S32) zum wiederholten Verändern des Zielsprachsignals (Svc(Ni)) in Stimmlage, bis ein Maximum (Pmax) der Wahrscheinlichkeiten (P) eine vorherbestimte Wahrscheinlichkeit (Pth) oder höher erreicht.

2. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 1, worin, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) kleiner als die vorherbestimmte Wahrscheinlichkeit (Pth) ist, die Stimmlagenveränderungsmittel (9, 11, 3, 5) Stimmlagenanpassungsmittel (S24, S34) beinhalten zum Erhöhen oder Erniedrigen des Zielsprachsignals (Svc(Ni))
auf der Basis des vorherbestimmten Grades (Ni).

3. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 2, weiterhin umfassend:

Speichermittel (3; S8) zum zeitweisen Speichern der eingehenden Kommandostimme (Svd);
Auslesensteuermittel (5, S4, S6, S8) zum Auslesen eines Strings der eingehenden Kommandostimme (Svd) aus den Speichermitteln (3) und Erzeugen des Zielsprachsignals (Svd); und
Auslesetaktsteuermittel (11, S22) zum Erzeugen eines Auslesetaktsignals (Scc) mit einem Zeittakt, der durch Frequenz bestimmt ist, und Ausgeben des Zeitaktes (Scc) an die Speichermittel (3), um mit dem dadurch spezifizierten Timing das Zielsprachsignal (Svc(Ni)) in Frequenz zu verändern auf Basis des vorherbestimmten Grades (Ni).

4. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 2, worin das Zielsprachsignal (Svc(Ni)) ausgehend von einem Lagenniveau der eingehenden Kommandostimme (Scd) in Stimmlage erhöht wird, basierend auf dem vorherbestimmten Grad (Ni).

5. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 4, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist bis zu einer ersten vorherbestimmten Lage (Mmax), und wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (Svc(Ni)) die erste vorherbestimmte Lage (Mmax) erreicht, das Zielsprachsignal (Svc(Ni)) in Stimmlage erniedrigt wird basierend auf dem vorherbestimmten Grad (Ni) ausgehend von dem Lagenniveau der eingehenden Kommandostimme (Svd).

6. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 5, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist nach unten auf eine zweite vorherbestimmte Lage (Mmin), und wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht, bevor das Zielsprachsignal (Svc(Ni)) die zweite vorherbestimmte Lage (Mmin) erreicht, die eingehende Kommandostimme angehalten wird, normalisiert zu werden.

7. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 2, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage erniedrigt wird auf den vorbestimmten Grad (Ni) ausgehend von einem Lagenniveau der eingehenden Kommandostimme (Scd).

8. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 7, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist nach unten auf eine dritte vorherbestimmte Lage (Mmin), und wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (Svc(Ni)) die dritte vorherbestimmte Lage (Mmin) erreicht, das Zielsprachsignal (Svc(Ni)) in Stimmlage erhöht wird, basierend auf dem vorherbestimmten Grad (Ni) ausgehend von dem Lagenniveau der eingehenden Kommandostimme (Svd).

9. Stimmlagen-Normalisierungsvorrichtung (Tr) nach Anspruch 8, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist nach unten auf eine vierte vorherbestimmte Lage (Mmax), und wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (Svc(Ni)) die vierte vorherbestimmte Lage (Mmax) erreicht, die eingehende Kommandostimme angehalten wird, normalisiert zu werden.

10. Spracherkennungsvorrichtung (VRAp) zum Erkennen einer eingehenden Kommandostimme (Sva), welche zur Spracherkennung optimal normalisiert ist basierend auf Musterdaten (Psf) für eine Vielzahl von Wörtern, wobei die Vorrichtung umfasst:

Zielspracherzeugungsmittel (9, 11, 3, 5; S24, S34, S26, S12) zum Erzeugen eines Zielprachsignals (Svc(Ni)) durch Verändern der eingehenden Kommandostimme (Svd) auf Basis eines vorherbestimmten Grades (Ni);
Wahrscheinlichkeitsberechnungsmittel (15; S14, #100) zum Berechnen einer Wahrscheinlichkeit (P), welche anzeigend für einen Grad der Übereinstimmung unter dem Zielsprachsignal (Svc(Ni)) und den Wörtern in den Musterdaten (Psf) ist; und
Stimmlagenveränderungsmittel (9, 11, 3, 5; S18, S20, S32) zum wiederholten Verändern des Zielsprachsignals (Svc(Ni)) in Stimmlage, bis ein Maximum (Pmax) der Wahrscheinlichkeiten (P) eine vorherbestimmte Wahrscheinlichkeit (Pth) oder höher erreicht.

11. Spracherkennungsvorrichtung (VRAp) nach Anspruch 10, worin, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) kleiner als die vorherbestimmte Wahrscheinlichkeit (Pth) ist, das Zielspracherzeugungsmittel Stimmlagenanpassungsmittel (S24, S34) beinhaltet zum Erhöhen oder Erniedrigen des Zielsprachsignals (Svc(Ni)) auf der Basis des vorherbestimmten Grades (Ni).

12. Spracherkennungsvorrichtung (VRAp) nach An-
spruch 11, weiterhin umfassend:

Speichermittel (3; S8) zum zeitweisen Speichern der eingehenden Kommandostimme (Svd);
Auslesesteuermittel (5, S4, S6, S8) zum Auslesen eines Strings der eingehenden Kommandostimme (Svd) aus den Speichermitteln (3) und Erzeugen des Zielsprachsignals (Svd); und
Auslesetaktsteuermittel (11, S22) zum Erzeugen eines Auslesetaktsignals (Scc) mit einem Zeittakt, der durch Frequenz bestimmt ist, und Ausgeben des Zeittaktes (Scc) an die Speichermittel (3), um mit dem dadurch spezifizierten Timing das Zielsprachsignal (SvCc(Ni)) in Frequenz zu verändern auf Basis des vorherbestimmten Grades (Ni).

13. Spracherkennungsvorrichtung (VRAp) nach Anspruch 11, worin das Zielsprachsignal (SvCc(Ni)) ausgehend von einem Lagenniveau der eingehenden Kommandostimme (Scd) in Stimmlage erhöht wird, basierend auf dem vorherbestimmten Grad (Ni).

14. Spracherkennungsvorrichtung (VRAp) nach Anspruch 13, worin das Zielsprachsignal (SvCc(Ni)) in Stimmlage begrenzt ist nach unten bis zu einer ersten vorherbestimmten Lage (Mmin), und, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (SvCc(Ni)) die erste vorherbestimmte Lage (Mmax) erreicht, das Zielsprachsignal (SvCc(Ni)) in Stimmlage erniedrigt wird basierend auf dem vorherbestimmten Grad (Ni) ausgehend von dem Lagenniveau der eingehenden Kommandostimme (Svd).

15. Spracherkennungsvorrichtung (VRAp) nach Anspruch 14, worin das Zielsprachsignal (SvCc(Ni)) in Stimmlage begrenzt ist nach unten bis zu einer zweiten vorherbestimmten Lage (Mmin), und, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (SvCc(Ni)) die zweite vorherbestimmte Lage (Mmin) erreicht, die eingehende Kommandostimme angehalten wird, normalisiert zu werden.

16. Spracherkennungsvorrichtung (VRAp) nach Anspruch 11, worin das Zielsprachsignal (SvCc(Ni)) in Stimmlage erniedrigt wird auf Basis des vorherbestimmten Grades (Ni) ausgehend von einem Lagenniveau der eingehenden Kommandostimme (Svd).

17. Spracherkennungsvorrichtung (VRAp) nach Anspruch 16, worin das Zielsprachsignal (SvCc(Ni)) in Stimmlage begrenzt ist nach unten auf eine dritte vorherbestimmte Lage (Mmin), und, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (SvCc(Ni)) die dritte vorherbestimmte Lage (Mmin) erreicht hat, das Zielsprachsignal (SvCc(Ni)) in Stimmlage erhöht wird, basierend auf dem vorherbestimmten Grad (Ni) ausgehend von dem Lagenniveau der eingehenden Kommandostimme (Svd).

18. Spracherkennungsvorrichtung (VRAp) nach Anspruch 17, worin das Zielsprachsignal (SvCc(Ni)) in Stimmlage begrenzt ist nach unten auf eine vierte vorherbestimmte Lage (Mmax), und, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (SvCc(Ni)) die vierte vorherbestimmte Lage (Mmax) erreicht, die eingehende Kommandostimme angehalten wird, normalisiert zu werden.

19. Stimmlagennormalisierungsverfahren, welches für eine Spracherkennungsvorrichtung (VRAp) verwendet wird zum Erkennen einer eingehenden Kommandostimme (Sva), welche durch irgendeinen Sprecher geäußert wird, basierend auf Musterdaten (Psf) für eine Vielzahl von Wörtern, und angewandt, um die eingehende Kommandostimme (SvC) zu normalisieren, so dass sie in einer optimalen Lage der Spracherkennung ist, wobei das Verfahren umfasst:

21. Stimmlagennormalisierungsverfahren nach Anspruch 20, weiterhin umfassend:

- einen Schritt (S8) des vorübergehenden Speichers der eingehenden Kommandostimme (Svd);
- einen Schritt (S4, S6, S8) des Erzeugens des Zielsprachsignals (Svd) von einem String der vorübergehend gespeicherten eingehenden Kommandostimme (Svd); und
- einen Schritt (S22) des Bestimmens eines Zeittales durch Frequenz, auf solche Weise, um mit dem dadurch spezifizierten Timing das Zielsprachsignal (Svc) in Frequenz zu verändern auf der Basis des vorherbestimmten Grades (Ni).

22. Stimmlagennormalisierungsverfahren nach Anspruch 20, weiterhin umfassend einen Schritt (S24) des Erhöhens des Zielsprachsignals (Svc(Ni)) in Stimmlage auf der Basis des vorherbestimmten Grades (Ni) ausgehend von einem Lagenniveau der eingehenden Kommandostimme (Svd).

23. Stimmlagennormalisierungsverfahren nach Anspruch 22, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist nach oben auf eine erste vorherbestimmte Lage (Mmax), und das Verfahren weiterhin einen Schritt (S30, S34) umfasst des, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (Svc(Ni)) erreicht, Erhöhens des Zielsprachsignals (Svc(Ni)) in Stimmlage auf der Basis des vorherbestimmten Grades (Ni) ausgehend von dem Lagenniveau der eingehenden Kommandostimme (Svd).

24. Stimmlagennormalisierungsverfahren nach Anspruch 23, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist nach unten auf eine zweite vorherbestimmte Lage (Mmin), und das Verfahren weiterhin einen Schritt (S32) umfasst des, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (Svc(Ni)) erreicht, Anhaltens des Normalisierens der eingehenden Kommandostimme.

25. Stimmlagennormalisierungsverfahren nach Anspruch 20, weiterhin umfassend einen Schritt (S34) des Ermiedrigens des Zielsprachsignals (Svc(Ni)) in Stimmlage auf der Basis des vorherbestimmten Grades (Ni) ausgehend von einem Lagenniveau der eingehenden Kommandostimme (Svd).

26. Stimmlagennormalisierungsverfahren nach Anspruch 25, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist nach unten auf eine dritte vorherbestimmte Lage (Mmin), und das Verfahren weiterhin einen Schritt (S24) umfasst des, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (Svc(Ni)) die dritte vorherbestimmte Lage (Mmin) erreicht, Erhöhens des Zielsprachsignals (Svc(Ni)) in Stimmlage auf der Basis des vorherbestimmten Grades (Ni) ausgehend von dem Lagenniveau der eingehenden Kommandostimme (Svd).

27. Stimmlagennormalisierungsverfahren nach Anspruch 26, worin das Zielsprachsignal (Svc(Ni)) in Stimmlage begrenzt ist nach unten auf eine vierte vorherbestimmte Lage (Mmax), und das Verfahren weiterhin einen Schritt (S32) umfasst des, wenn das Maximum (Pmax) der Wahrscheinlichkeiten (P) die vorherbestimmte Wahrscheinlichkeit (Pth) oder höher nicht erreicht hat, bevor das Zielsprachsignal (Svc(Ni)) erreicht, Erhöhens des Zielsprachsignals (Svc(Ni)) die vierte vorherbestimmte Lage (Mmax) erreicht, Anhaltens des Normalisierens der eingehenden Kommandostimme.

Revendications

1. Dispositif (Tr) de normalisation de la hauteur tonale de la voix monté dans un dispositif de reconnaissance de la parole (VRAp), destiné à identifier une commande vocale incidente (Sva) émise par un locuteur quelconque, en se basant sur des données échantillonnées (Psf) pour une pluralité de mots, et utilisé pour normaliser la commande vocale incidente (Svc) de manière à ce qu'elle soit dans une hauteur tonale optimale pour la reconnaissance de la parole, le dispositif comprenant :

   un moyen (9, 11, 3, 5 ; S24, S34, S26, S12) de génération d’un signal vocal cible, destiné à générer un signal vocal cible (Svc(Ni)) en modifiant ladite commande vocale incidente (Svd) en se basant sur un degré prédéterminé (Ni) ; un moyen (15 ; S14, #100) de calcul de probabilité, destiné à calculer une probabilité (P) indiquant un degré de coïncidence entre l’identité vocale (Svd) et les mots dans les dites données échantillonnées (Psf) ; et un moyen (9, 11, 3, 5 ; S18, S20, S32) de modification de la hauteur tonale de la voix, destiné à modifier de façon répétée l’identité vocale cible (Svc(Ni)) en ce qui concerne la hauteur tonale de la voix jusqu'à ce qu’un maximum (Pmax) desdites probabilités (P) atteigne une probabilité prédéterminée (Pth) ou une proba-
6. Dispositif (Tr) de normalisation de la hauteur tonale de la voix selon la revendication 5, dans lequel le signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une deuxième hauteur tonale prédéterminée (Mmin), et si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la deuxième hauteur tonale prédéterminée (Mmin), la normalisation de ladite commande vocale incidente est arrêtée.

7. Dispositif (Tr) de normalisation de la hauteur tonale de la voix selon la revendication 2, dans lequel le signal vocal cible (Svc(Ni)) est diminué en ce qui concerne la hauteur tonale de la voix en se basant sur ledit degré prédéterminé (Ni), en partant d’un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

8. Dispositif (Tr) de normalisation de la hauteur tonale de la voix selon la revendication 7, dans lequel le signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une troisième hauteur tonale prédéterminée (Mmin), et si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la troisième hauteur tonale prédéterminée (Mmin), le signal vocal cible (Svc(Ni)) est augmenté en ce qui concerne la hauteur tonale en se basant sur ledit degré prédéterminé (Ni), en partant d’un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

9. Dispositif (Tr) de normalisation de la hauteur tonale de la voix selon la revendication 8, dans lequel le signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une quatrième hauteur tonale prédéterminée (Mmax), et si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la quatrième hauteur tonale prédéterminée (Mmax), la normalisation de ladite commande vocale incidente est arrêtée.

10. Dispositif (VRAp) de reconnaissance de la parole destiné à identifier une commande vocale incidente (Sva) normalisée de manière optimale pour la reconnaissance de la parole en se basant sur des données échantillonnées (PsS) pour une pluralité de mots, le dispositif comprenant :

   un moyen (9, 11, 3, 5 ; S24, S34, S26, S12) de...
génération de voix cible, destiné à générer un signal vocal cible (Svc(Ni)) en modifiant ladite commande vocale incidente (Svd) en se basant sur un degré prédéterminé (Ni) ; un moyen (15 ; S14, #100) de calcul de probabilité, destiné à calculer une probabilité (P) indiquant un degré de coincidence entre le signal vocal cible (Svc(Ni)) et les mots dans les données échantillonnées (Psf) ; et un moyen (9, 11, 3, 5 ; S18, S20, S32) de modification de la hauteur tonale de la voix, destiné à modifier de façon répétée le signal vocal cible (Svc(Ni)) en ce qui concerne la hauteur tonale de la voix jusqu’à ce qu’un maximum (Pmax) desdites probabilités (P) atteigne une probabilité prédéterminée (Pth) ou une probabilité supérieure.

11. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 10, dans lequel le maximum desdites probabilités (P) est inférieur à ladite probabilité prédéterminée (Pth), ledit moyen de modification de la hauteur tonale de la voix comprend un moyen (S24, S34) de réglage de la hauteur tonale de la voix destiné à augmenter ou à diminuer ledit signal vocal cible (Svc(Ni)) en se basant sur le degré prédéterminé (Ni).

12. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 11, comprenant en outre :

un moyen de mémoire (3, S8) destiné à mémoire temporairement ladite commande vocale incidente (Svd) ;
un moyen (5, S4, S6, S8) de commande de lecture destiné à lire une chaîne de ladite commande vocale incidente (Svd) à partir dudit moyen de mémoire (3), et à générer le signal vocal cible (Svc(Ni)) ; et
un moyen (11, S22) de commande d’horloge de lecture destiné à générer un signal (Scc) d’horloge de lecture au moyen d’une horloge de synchronisation commandée par la fréquence, et à délivrer ledit signal d’horloge (Scc) audit moyen de mémoire (3) de manière à modifier en fréquence, avec la synchronisation spécifiée à cet effet, ledit signal vocal cible (Svc(Ni)) en se basant sur ledit degré prédéterminé (Ni).

13. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 11, dans lequel le signal vocal cible (Svc(Ni)) est augmenté en ce qui concerne la hauteur tonale de la voix en se basant sur ledit degré prédéterminé (Ni), en partant d’un niveau de hauteur tonale de ladite commande vocale incidente (Svd).

14. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 13, dans lequel le signal vocal cible (Svc(Ni)) est limité vers le haut en ce qui concerne la hauteur tonale de la voix à une première hauteur tonale prédéterminée (Mmax), et si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la première hauteur tonale prédéterminée (Mmax), ledit signal vocal cible (Svc(Ni)) est diminué en ce qui concerne la hauteur tonale en se basant sur ledit degré prédéterminé (Ni), en partant d’un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

15. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 14, dans lequel le signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une deuxième hauteur tonale prédéterminée (Mmin), et si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la deuxième hauteur tonale prédéterminée (Mmin), la normalisation de ladite commande vocale incidente est arrêtée.

16. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 11, dans lequel le signal vocal cible (Svc(Ni)) est diminué en ce qui concerne la hauteur tonale de la voix en se basant sur un degré prédéterminé (Ni), en partant d’un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

17. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 16, dans lequel le signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une troisième hauteur tonale prédéterminée (Mmin), et si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la troisième hauteur tonale prédéterminée (Mmin), ledit signal vocal cible (Svc(Ni)) est augmenté en ce qui concerne la hauteur tonale en se basant sur ledit degré prédéterminé (Ni), en partant d’un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

18. Dispositif (VRAp) de reconnaissance de la parole selon la revendication 17, dans lequel le signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une quatrième hauteur tonale prédéterminée (Mmax), et si ledit maximum (Pmax) desdites probabilités (P) ne par-
vient à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n'atteigne la quatrième hauteur tonale prédéterminée (Mmax), la normalisation de ladite commande vocale incidente est arrêtée.

19. Procédé de normalisation de la hauteur tonale de la voix utilisé dans un dispositif (VRAp) de reconnaissance de la parole destiné à identifier une commande vocale incidente (Sva) émise par un locuteur quelconque, en se basant sur des données échantillonnées (Psf) pour une pluralité de mots, et utilisé pour normaliser la commande vocale incidente (Svc) de manière à ce qu'elle soit dans une hauteur tonale optimale pour la reconnaissance de la parole, le procédé comprenant :

une étape (S24, S34, S26, S12) consistant à générer un signal de voix cible en modifiant la dite commande vocale incidente (Svd) en se basant sur un degré prédéterminé (Ni) ;
une étape (S14, #100) consistant à calculer une probabilité (P) indiquant un degré de coïncidence entre ledit signal vocal cible (Svc(Ni)) et les mots dans lesdites données échantillonnées (Psf) ; et
une étape (S18, S20, S32) consistant à modifier de façon répétée ledit signal vocal cible (Svc(Ni)) en ce qui concerne la hauteur tonale de la voix jusqu'à ce qu'un maximum (Pmax) desdites probabilités (P) atteigne une probabilité supérieure.

20. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 19, comprenant en outre une étape (S24, S34) consistant, si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n'atteigne la première hauteur tonale prédéterminée (Mmax), à diminuer ledit signal vocal cible (Svc(Ni)) en ce qui concerne la hauteur tonale en se basant sur ledit degré prédéterminé (Ni), en partant d'un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

21. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 20, comprenant en outre :

une étape (S8) consistant à mémoriser temporairement ladite commande vocale incidente (Svd) ;
une étape (S4, S6, S8) consistant à générer le signal vocal cible (Svc(Ni)) à partir d'une chaîne de ladite commande vocale incidente (Svd) mémorisée temporairement ; et
une étape (S22) consistant un signal de synchronisation par la fréquence, d'une manière telle qu'elle modifie, avec la synchronisation spécifiée à cet effet, ledit signal vocal cible (Svc) en se basant sur ledit degré prédéterminé (Ni).

22. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 20, comprenant en outre une étape (S24) consistant à augmenter ledit signal vocal cible (Svc(Ni)) en ce qui concerne la hauteur tonale de la voix en se basant sur ledit degré prédéterminé (Ni), à partir d'un niveau de hauteur vocale de ladite commande vocale incidente (Svd).

23. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 22, dans lequel ledit signal vocal cible (Svc(Ni)) est limité vers le haut en ce qui concerne la hauteur tonale de la voix à une première hauteur tonale prédéterminée (Mmax), et le procédé comprend en outre une étape (S30, S34) consistant, si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n'atteigne la première hauteur tonale prédéterminée (Mmax), à diminuer ledit signal vocal cible (Svc(Ni)) en ce qui concerne la hauteur tonale en se basant sur ledit degré prédéterminé (Ni), en partant d'un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

24. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 23, dans lequel ledit signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une deuxième hauteur tonale prédéterminée (Mmin), et le procédé comprend en outre une étape S32 consistant, si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n'atteigne la deuxième hauteur tonale prédéterminée (Mmin), à arrêter la normalisation de ladite commande vocale incidente.

25. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 20, comprenant en outre une étape (S34) consistant à diminuer ledit signal vocal cible (Svc(Ni)) en ce qui concerne la hauteur tonale de la voix en se basant sur ledit degré prédéterminé (Ni), en partant d'un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

26. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 25, dans lequel ledit signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une troisième hauteur tonale prédéterminée (Mmin), et le procédé comprend en outre une étape
(S24) consistant, si ledit maximum (Pmax) desdites probabilités (P) ne parvient à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la troisième hauteur tonale prédéterminée (Mmin), à augmenter ledit signal vocal cible (Svc (Ni)) en ce qui concerne la hauteur tonale en se basant sur ledit degré prédéterminé (Ni), en partant d’un niveau de hauteur tonale dudit signal de commande vocale incidente (Svd).

27. Procédé de normalisation de la hauteur tonale de la voix selon la revendication 26, dans lequel ledit signal vocal cible (Svc(Ni)) est limité vers le bas en ce qui concerne la hauteur tonale de la voix à une quatrième hauteur tonale prédéterminée (Mmax), et

le procédé comprend en outre une étape (S32) consistant, si ledit maximum (Pmax) desdites probabilités (P) ne parvient pas à atteindre ladite probabilité prédéterminée (Pth) ou une probabilité supérieure avant que le signal vocal cible (Svc(Ni)) n’atteigne la quatrième hauteur tonale prédéterminée (Mmax), à arrêter la normalisation de ladite commande vocale incidente.
Fig. 3

Fig. 4
Fig. 6

Start

S102 Obtain PSvc(Ni)

S104

m = 1,
Pmax(Ni) = 0
Sr' = 0

S106 Obtain Psf(m), Sr(m)

S108 Calculate P(m)

S110

P(m) ≥ Pmax ?

No

Yes

S112 Pmax(Ni) = P(m)

S114 Produce Csr

S116 Sr' = Sr(m)

S118 m = M ?

No

Yes

S122 Output Pmax(Ni)

End

S120 m = m + 1