

TITLE OF THE INVENTION

COARTICULATION PROCESSING APPARATUS FOR CHINESE
SPEECH SYNTHESIS

BACKGROUND OF THE INVENTION

5 1. FIELD OF THE INVENTION

The present invention relates to a coarticulation processing apparatus for Chinese speech synthesis, and in particular, a coarticulation processing apparatus for obtaining a smooth transition from one syllable to a succeeding one syllable in Chinese speech synthesis.

2. DESCRIPTION OF THE RELATED ART

In Chinese speech synthesis, a process for smoothing the connection of adjacent syllables in order to make the pronunciation of a series of syllables more fluent and smooth is called coarticulation process. In order to obtain a smooth transition from one of the syllables of a word string to a succeeding one of the syllables of the word string in Chinese speech synthesis, a coarticulation process that overlaps a part of the phonemes constituting the preceding syllable with part of the phonemes constituting the succeeding syllable is required.

Fig. 3 is a broadband spectrogram of an utterance "中文" (Chinese language) that is produced by a human. It is apparent from Fig. 3 that there is a coarticulation phenomenon. However, most of the conventional Chinese speech synthesizing

systems simply interconnect the neighboring phonemes of the two syllables of the word string "中文", without taking the coarticulation process into consideration, as shown in Fig. 4.

Thus, this results in a synthesized speech whose naturalness
5 is low.

Furthermore, the coarticulation processing techniques employed in the conventional Chinese speech synthesizing systems are to perform the simulation in time domain on the coarticulation segment of the word string, that is, an
10 optimum coarticulation segment is first of all searched from a large number of recorded coarticulation segment speech data.

Then, the optimum coarticulation segment is interpolated between the preceding syllable and the succeeding syllable. The critical points of the above process are the determination of
15 the optimum coarticulation segment and the search for the optimum coarticulation segment from the recorded coarticulation segment speech data. The paper, entitled "中文連音二字詞之語音合成" (Chinese speech synthesis of coarticulation two-character word), presented to the Republic Of China, Computational Linguistics
20 Conference IX (1996), is incorporated herein for reference.

Fig. 5 is a schematic system block diagram of the aforementioned conventional Chinese speech synthesizing system.

Referring to Fig. 5, 100 denotes a phonetic
25 transcribing sentence input unit which enables an operator to

input a phonetic transcribing sentence that is to be synthesized.

110 denotes a word string storage unit for storing a large number of recorded word string speech data. 180 denotes a monosyllable storage unit for storing recorded monosyllable speech data. 120

5 denotes a word string searching unit for searching from the word string storage unit 110 for a word string that has to be undergone a coarticulation process in accordance with the input phonetic transcribing sentence, and parses the searched word string to determine the coarticulation segment. 130 denotes a central

10 position searching unit for searching for the central position of the coarticulation segment of the word string. 140 denotes an evaluation unit for evaluating the phonemic time length of the coarticulation segment. 150 denotes a front syllable synthesizing unit for searching the monosyllable storage unit

15 180 for the recorded monosyllable speech data of the preceding syllable in accordance with the input phonetic transcribing sentence and for synthesizing the searched recorded monosyllable speech data. 160 denotes a coarticulation segment synthesizing unit for synthesizing the synthesized speech data

20 output from the front syllable synthesizing unit 150 with the coarticulation segment. 170 denotes a back syllable synthesizing unit for searching the monosyllable storage unit

25 180 for the recorded monosyllable speech data of the succeeding syllable in accordance with the input phonetic transcribing sentence and for synthesizing the synthesized speech data output

from the coarticulation segment synthesizing unit 160 with the searched recorded monosyllable speech data of the succeeding syllable. 190 denotes a synthesized speech output unit for outputting the resultant synthesized speech data output in speech.

It is apparent from Fig. 5 that the aforementioned conventional Chinese speech synthesizing system searches for the optimum coarticulation segment from the word string storage unit 110 and the recorded monosyllable speech data from the monosyllable storage portion 180, and synthesizes them in order to improve the naturalness and comprehensibility of the synthesized speech output.

For example, if the word string "中文" (Chinese language), which is a word string that is to be undergone the coarticulation process, is to be synthesized by using the system shown in Fig. 5, the phonetic transcribing sentence corresponding to the word string "中文" is first of all input by an operator through the input unit 100. Afterwards, the word string storage unit 110 is searched for a recorded word string speech data corresponding to the word string "中文". Supposing that there has the recorded word string speech data of the word string "中文" in the word string storage unit 110, the word string searching unit 120 searches the recorded word string speech data corresponding to the word string "中文" from the word string storage unit 110. The searched recorded word string speech data

is parsed so as to determine the coarticulation segment of the word string "中文". The central position of the coarticulation segment of the word string "中文" is evaluated in the central position searching unit 130. The phonemic time length of the coarticulation segment of the word string "中文" is evaluated in the evaluation unit 140. The front syllable synthesizing unit 150 searches the monosyllable storage portion 180 for a recorded monosyllable speech data corresponding to the word "中". The coarticulation segment synthesizing unit 160 synthesizes the searched monosyllable recorded speech data corresponding to the word "中" with the coarticulation segment. Then, the back syllable synthesizing unit 170 searches the monosyllable storage unit 180 for a recorded monosyllable speech data corresponding to the word "文" and synthesizes the synthesized speech data output from the coarticulation segment synthesizing unit 160 with the searched recorded monosyllable speech data corresponding to the word "文". Finally, the synthesized speech output unit 190 outputs the resultant synthesized speech data output in speech.

However, if the word string storage unit 110 does not contain the recorded word string speech data corresponding to the word string "中文", a most similar coarticulation segment will be determined in accordance with the vowel (ㄛ) of the preceding syllable of the word string "中文" and the initial phoneme (ㄌ) of the succeeding syllable of the word string

"中文", for example, "通問" (トウモン メウ), and is synthesized as described above. Thus, this results in a synthesized speech whose naturalness is low. Furthermore, the aforementioned system requires a memory space of about 55 M Bytes for storing a large number of recorded word string speech data, thereby consuming the valuable memory space. Additionally, since recorded speech data are used as the basic unit for synthesis, the frequency and phonemic time length cannot be varied, and the searching and synthesizing of the recorded speech data is time-consuming.

Accordingly, the aforementioned prior art has the following drawbacks:

1. It is necessary to store a large number of recorded monosyllable speech data and recorded word string speech data.

2. It results in synthesized speech whose naturalness is low if the recorded word string speech data that is desired is not contained in the word string storage unit.

3. Since recorded speech data are used, the phonemic time length and the rhythm cannot be varied.

4. The searching of the recorded speech data is time-consuming.

SUMMARY OF THE INVENTION

Therefore, an essential object of the present invention is to provide a coarticulation processing apparatus for obtaining a smooth transition from one syllable to a

succeeding one syllable in Chinese speech synthesis, which coarticulation processing apparatus can overcome the drawbacks that are associated with the aforementioned prior art.

According to the present invention, a coarticulation
5 processing apparatus for Chinese speech synthesis, comprising:

a dictionary memory for storing a dictionary including a plurality of Chinese word strings and the corresponding phonetic transcriptions therein;

10 a storage unit for storing pitch data of various Chinese syllables and coarticulation segments and the phonetic transcriptions corresponding to said various Chinese syllables and coarticulation segments, and starting and ending points of the consonants and vowels of said various Chinese syllables and coarticulation segments;

15 a word parsing unit for parsing an input phonetic transcribing sentence to be synthesized based on the dictionary stored in said dictionary memory to section the sentence into a plurality of word strings;

20 a syllable parsing unit for determining which word string from said word parsing unit should be undergone a coarticulation process based on said storage unit so as to search coarticulation segment of determined said word string that should be undergone the coarticulation process; and

25 a speech synthesizing unit for interpolating searched said coarticulation segment between said syllables of said word

string in said input phonetic transcribing sentence and generating a synthesized speech.

5 In the above-mentioned coarticulation processing apparatus, the storage unit preferably stores 409 Chinese syllables with Chinese first tone.

In the above-mentioned coarticulation processing apparatus, the coarticulation segments stored in the storage unit are initial phonemes of succeeding syllables of Chinese word strings, for example, as defined in Fig. 6.

10 The CV-VC(VV) coarticulation processing apparatus for Chinese speech synthesis of the present invention having the above-described construction first of all sections the phonetic transcribing sentence input by a user into a plurality of word strings based on the dictionary stored in the dictionary
15 memory. Then, the syllable parsing unit determines the preceding and succeeding syllables that have to be undergone a coarticulation process. Afterwards, the pitch data and the starting and ending point of the consonant and vowel of each syllable are searched from the syllable data storage unit.
20 Finally, the speech synthesizing unit evaluates the phonemic time length and frequency for the variation operation of the phonemic time length and frequency, and synthesizes and outputs the speech.

BRIEF DESCRIPTION OF THE DRAWINGS

25 These and other objects and features of the present

invention will become clear from the following description taken in conjunction with the preferred embodiments thereof with reference to the accompanying drawings throughout which like parts are designated by like reference numerals, and in which:

5 Fig. 1 is a schematic system block diagram of a coarticulation processing apparatus for use in Chinese speech synthesis of a preferred embodiment according to the present invention;

10 Fig. 2 is a diagram illustrating the contents of syllables stored in a register unit 13 shown in Fig. 1;

 Fig. 3 is a broadband spectrogram of an utterance "中文" (Chinese language) that is produced by a human;

15 Fig. 4 is a broadband spectrogram of an utterance "中文" (Chinese language) that is produced by a conventional Chinese speech synthesizing system;

 Fig. 5 is a schematic system block diagram of a conventional Chinese speech synthesizing system;

20 Fig. 6 is a diagram illustrating the type of initial phonemes of succeeding syllables that are used for determining whether or not a word string is to be undergone a coarticulation process in the preferred embodiment; and

 Fig. 7 is a diagram for explaining a coarticulation process for the word string "台灣" (Taiwan) in the preferred embodiment according to the present invention.

25 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Preferred embodiments according to the present invention will be described below with reference to the attached drawings.

Fig. 1 is a schematic system block diagram of a
5 coarticulation processing apparatus for Chinese speech synthesis of a preferred embodiment according to the present invention.

Referring to Fig. 1, an input unit 10 comprises, for example, a keyboard, which enables an operator to input a
10 phonetic transcribing sentence that is to be synthesized in speech. A word parsing unit 11 parses the input sentence based on a dictionary stored in a dictionary memory (storage unit) 12 in which a plurality of word strings and the corresponding phonetic transcription are stored to section the sentence into
15 a plurality of word strings and to label the position between two adjacent word strings. A syllable parsing unit 14 determines which word string should be undergone a coarticulation process based on the pitch data storage unit 15 of VC(VV) coarticulation segment and CV syllable and the label
20 data storage unit 16 of VC(VV) coarticulation segment and CV syllable, and searches for the pitch data and label data of the determined CV syllable and VC(VV) coarticulation segment. A syllable time length searching unit 17 and a syllable frequency
25 searching unit 18 searches the corresponding time length and frequency of the syllable based on the rhythm rule. It is noted

that C means consonant, and V means vowel.

A register (storage) unit 13 stores the corresponding time length, frequency, tone (which is Chinese tone, that is an accent attached to a Chinese syllable), and phonetic symbol of each syllable therein. A waveform overlapping and adding unit 19 overlaps and adds the waveforms of the CV syllable and VC(VV) coarticulation segment. A synthesized speech output unit 20 outputs the resultant synthesized speech.

Application of the present invention will be explained as follows. For example, in a phonetic transcribing sentence,

[tai2 wan1 shi4 yi2 ge5 mei3 li4 de5 bao3 dao3

(台灣是一個美麗的寶島)]

(English translation: Taiwan is one beautiful treasure island), the phonetic transcription of each of the words is followed by a number that represents the tone of the corresponding word.

Fig. 6 is a diagram illustrating the type of the initial phonemes of the succeeding syllables that are used for determining whether or not a word string is to be undergone a coarticulation process in the preferred embodiment.

The coarticulation process of the preferred embodiment according to the present invention is performed as follows: the phonetic transcribing sentence as described above is first of all input by the operator using the input unit 10.

The word parsing unit 11 parses the input sentence based on

the dictionary stored in the dictionary memory 12 to section the sentence into a plurality of word strings and to label the position between two adjacent word strings, so that a phonetic transcribing list containing label data,

5 tai2 wan1 @ shi4 @ yi2 ge5 @ mei3 li4 @ de5 @ bao3 dao3,
is generated, where @ is the label data for labeling the position between two adjacent word strings. Then, the syllable parsing unit 14 determines which word string should be undergone a coarticulation process in accordance with the phonemes shown
10 in Fig. 6.

As a result, it is found that the phonemes in the word strings,

"tai2 wan1" and "mei3 li4",
are needed to undergo the coarticulation process. The syllable
15 parsing unit 14 searches the pitch data storage unit 15 of VC(VV) coarticulation segment and CV syllable, and the label data storage unit 16 of VC(VV) coarticulation segment and CV syllable, for the pitch data and label data of the VC(VV) coarticulation segment and the CV syllable in accordance with the table shown
20 in Fig. 6, and then, stores the pitch data and the label data into the register unit 13. That is,

tai2 aiwan wan1 @ shi4 @ yi2 ge5 @ mei3 eil li4 @ de5 @
bao3 dao3,

are searched from the storage units 15 and 16, wherein aiwan
25 (アイワン) is a VV coarticulation segment and eil (エイ) is a VC

coarticulation segment. The syllable time length searching unit 17 and the syllable frequency searching unit 18 determines and stores the corresponding time length and frequency of each syllable into the register unit 13.

5 Fig. 2 is a diagram illustrating the contents of syllables stored in a register unit 13 shown in Fig. 1. In the preferred embodiment, the register unit 13 stores 409 Chinese syllables with Chinese first tone among Chinese four tones.

10 The waveform overlapping and adding unit 19 overlaps and adds the waveforms of the CV syllable and VC(VV) coarticulation segment based on the detail data of the CV syllable and VC(VV) coarticulation segment from the register 13, including the time length of the syllable, the time length of the consonant of the syllable, the starting point of the syllable, the ending point of the syllable, the frequency of
15 the eight sections of the syllable, the type of the tone of the syllable, the type of the consonant of the syllable, the type of the vowel of the syllable, the position of the syllable in the word string, the serial number of the CV syllable of the
20 syllable, the serial number of the VC(VV) coarticulation segment of the syllable, etc. Finally, the synthesized speech output unit 20 outputs the resultant synthesized speech..

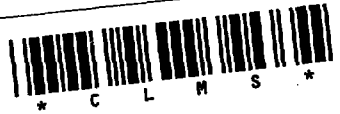
 Fig. 7 is a diagram for explaining the coarticulation process for a word string, "台灣" (Taiwan) in the preferred
25 embodiment. First of all, the pitch data and the label data of

each CV syllable and its VV coarticulation segment are stored in the register unit 13. Then, the frequency and the time length of the syllables "台" and "灣", and the coarticulation segment "ㄟㄨㄢ" are evaluated based on the rhythm rule so as to overlap
5 and add the waveform of the syllables and the coarticulation segment in order to generate the waveform of the word string "台灣". Since the synthesis of the word string "台灣" is accomplished with the pitch data, so that the time length and frequency can be varied, and the valuable memory space can be
10 saved.

As has been described hereinbefore, the preferred embodiment of the present invention overcomes the problems in coarticulation that are associated with the prior art. By searching the coarticulation segment of a word string and by
15 overlapping and adding the waveforms of the coarticulation segment with the preceding syllable and the succeeding syllable of the word string, a synthesized speech whose naturalness is high can be obtained. Furthermore, the time length and frequency of the word string can be varied so that word strings
20 with different tone and time length can be produced and the valuable memory space can be saved.

Although the present invention has been fully described in connection with the preferred embodiments thereof with reference to the accompanying drawings, it is to be noted that
25 various changes and modifications are apparent to those skilled

in the art. Such changes and modifications are to be understood as included within the scope of the present invention as defined by the appended claims unless they depart therefrom.



WHAT IS CLAIMED IS:

1. Acoarticulation processing apparatus for Chinese speech synthesis, comprising:

5 a dictionary memory for storing a dictionary including a plurality of Chinese word strings and the corresponding phonetic transcriptions therein;

10 a storage unit for storing pitch data of various Chinese syllables and coarticulation segments and the phonetic transcriptions corresponding to said various Chinese syllables and coarticulation segments, and starting and ending points of the consonants and vowels of said various Chinese syllables and coarticulation segments;

15 a word parsing unit for parsing an input phonetic transcribing sentence to be synthesized based on said dictionary stored in said dictionary memory, to section the sentence into a plurality of word strings;

20 a syllable parsing unit for determining which word string from said word parsing unit should be undergone a coarticulation process based on said storage unit so as to search coarticulation segment of determined said word string that should be undergone the coarticulation process; and

25 a speech synthesizing unit for interpolating searched said coarticulation segment between said syllables of said word string in said input phonetic transcribing sentence and generating a synthesized speech.



2. A coarticulation processing apparatus as claimed
in claim 1,

wherein said storage unit stores 409 Chinese
syllables with Chinese first tone.

5 3. A coarticulation processing apparatus as claimed
in claim 1,

wherein said coarticulation segments stored in said
storage unit are the initial phonemes of the succeeding syllables
of Chinese word strings.

10 4. A coarticulation processing apparatus as claimed
in claim 2,

wherein said coarticulation segments stored in said
storage unit are the initial phonemes of the succeeding syllables
of Chinese word strings.