**S15.10**

**文本自动发音翻译**

K.M.·阿卜杜勒·阿齐兹

英国诺丁汉大学电机电子工程系

摘要

自动翻译文本语音是最有用的,当不局限于一个现有的词汇。

潜在的质量有明显的改进语音输出复杂等演讲芯片将有更多的控制参数。

这个质量改善需要更大的计算资源处理等过程中作为文本正常化和一种应用音位变体字符串的算法。

许多的过程都是理想的适于做并发运行。

一个multi-transputer系统建立了奥坎:安排上,完成该文本自动语音转换。

摘要

The automatic translation of text to speech is most useful when not restricted to a pre-stored vocabulary.

自动翻译文本语音是最有用的,当不局限于一个pre-stored词汇。

The potential quality of speech output has improved dramatically as the complexity of speech synthesis chips incorporate a greater number of control parameters.

潜在的质量有明显的改进语音输出复杂等演讲芯片将有更多的控制参数。

This improvement in quality necessitates greater computing resources for handling such processes as text normalisation and the application of an allophone string generation algorithm.

这个质量改善需要更大的计算资源处理等过程中作为文本正常化和一种应用音位变体字符串的算法。

Many of the processes involved are ideally suited to be run concurrently.

许多的过程都是理想的适于做并发运行。

A multi-transputer systerm programmed in Occam has been constructed to perform this automatic text to speech conversion.

一个multi-transputer系统建立了奥坎:安排上,完成该文本自动语音转换。

INTRODUCTION

论文简介

In the past,the simplest approach to unrestricted text to speech translation used a small set of letter to word rules 11.2.31 each specifyihg a pronunciation for one or more letters in any one context.

在过去,最简单的方法来限制使用文本语音翻译一小组给词的发音规则specifyihg 11.2.31每一个或多个字符在任何一个背景。

Unless this approach yielded sufficient intelligibility,the routine addition of text to speech translation to computer systems was unlikely since more elaborate approaches embodying large pronunciation dictionaries or linguistics use too much of the available sequential computing resources [4].

除非这个方法取得足够的清晰度,常务增加文本语音翻译的计算机系统不太可能更详细的方法从大的发音字典或语言学体现使用太多的可获得连续计算资源[4]。

The recent introduction of complex speech synthesis chip sets incorporating such features as stress, speed, pitch, amplitude and prosody information has further compounded the processing problems involved in the translation.

最近引入复杂的语音合成芯片集合把这样的特征,压力大、速度、沥青、振幅和韵律信息处理问题不可开交参与翻译。

A typical method for performing text to speech conversion requires several computational stages as

is shown in Figure 1.

为执行典型方法文本语音转换需要几个计算阶段

如图1所示。

The text normalisation stage converts the text into a standard form of text suitable for further processing.

文本正常化把文本转换阶段的标准形式的文本适合进一步的加工。

Thir is achieved by expanding abreviations, handling punctuations and altering non-alphabetic characters. Another stage is required for the generation of a phoneme string for the pronunciation of words that are exceptions to the pronunciation rule. These words require the use of a special dictionary. For words that have a standard pronunciation a stage is required for the application of a phoneme string generation algorithm using context sensitive rules.

以达到扩大abreviations同名,处理的符号,改变去非字母字符。需要另一个阶段生成线音位的字的读音,是例外的发音规则。这些话需要使用一种特殊的字典。说不出话来,有一个标准发音阶段中的应用要求音位的字符串生成算法利用上下文敏感的规则。

The phoneme strings producted may require further computation to interpret the correct versions of each phoneme using syntactic information and stress patterns for conversion into allophones.

弦音位生产计算可能需要进一步研究历史时,那些解释正确的版本的每个音素摘要采用句法信息和压力模式转变成了变体的。

The parameter generator then converts the selected sounds into the speech parameters for the synthesiser to create speech.

参数发生器然后转换选定的语音转换成语音合成器参数创建的演说。

The standard pronunciation algorithm stage comprises several processes.The recognition algorithm is based around one character within a character string. This character string can vary from one to three characters long,also associated with each string is a suffix and a prefix character. The characters can be letters, other characters, voiced consonants,vowel cluster etc. The processes involved in the pronunciation algorithm initially requires the identification of a letter character.Associated with each letter character is a set of rules. Once the correct rule set has been identified, then the specific rule from within the rule set must be found. This is achieved by studying the character string lengths and matching the prefix character, second character,third character and suffix character to the entries within the rule set. Associated with each rule entry is a string of allophones, thus when the relevant rule has been found from within the rule set, the correct allophone string can be passed to the sound section.

标准读音算法由一些阶段过程。识别算法是基于一个字符在一个字符串。这个字符串将不同三个大字长,也伴随每一个字符串是一个后缀和前缀的性格。字符就被信件,其它的字符,表示辅音、元音聚类等。所涉及的过程,在发音算法的识别最初要求信的性格。每个字母的性格是相关的一套规则。一旦正确的规则集已经确认,那麽特定的规则从内规则集必须被发现。达到这个目标通过研究匹配字符串长度和前缀而言,第二而言,第三的个性和后缀字符参赛规则集内。伴随每一个规则条目是一连串的变体,因此,当相关的规则已经被发现的规则集内,正确的音位变体字符串可以被传递给声音部分。

PARALLELISM IN THE TRANSLATION

Many of the processes involved in the automatic translation of text to speech, such as text normalisation, exception to the pronunciation NIC phoneme string generation, standard pronuniciation algorithm and prosody applications can be carried out concurrently. Indeed the

application of the processes involved within the pronunciation algorithm on individual elements within the text stream can be carried out concurrently. This concurrent nature of the processing involved in the automatic translation of text to speech makes it ideally suited to be programmed in Occam and run upon a transputer system [5.6.7]

许多所涉及的过程,在自动翻译文本语音,如文字正常化,异常对语音网卡音素字符串生成、标准pronuniciation算法和韵律申请可同时进行。的确

应用在发音过程每个元素的算法在文本流都可以进行并发执行。这并行性质的加工参与的自动翻译文本语音让这本书适合被编程在奥坎:和运行在一个transputer系统[5.6.7]

Single Transputer System

It was decided to investigate the concurrent nature of the standard pronunciation algorithm and its mapping onto a transputer system.The algorithm investigated is based on rules developed around the 50,ooO word standard corpus of present day edited American English [SI.The initial system developed is shown in Figure 2. The system is based upon a BOO4 transputer development board situated within an IBM compatible host computer. One of the BOO4 links is utilised to interface via a COO11 to the speech synthesis processor and audio amplifier.The SpoZ56 allow for the selection of 64 allophones and incorporates a digital filter that models the vocal tract.

单Transputer系统

这是决定调查并行性质的发音标准算法和其映射到一个transputer系统。该算法是基于规则进行回绕50如何词语料库今天的《标准美国英语[斯。最初的系统开发的是如图2所示。该系统是根据一BOO4位于transputer开发板在一个兼容IBM的主计算机中。BOO4链接就的介面,通过一个COO11用以话语合成处理器和音频放大器。SpoZ56允许选择的64位数字滤波,耐变体声带共振引起这个模型。

The programming of the system was achieved using Occam 2 (β+). The program is modular and menu driven from the point of view of easy expansion,good programming practice, care of use and care of testing. A flow diagram of the software is shown in Figure 3. The fmt module of the program performs the user interface. Another module strips out non-valid.

编程系统的β用到奥坎:2(以上)。这个计划是模块化和菜单驱动从的角度来看,容易扩展,良好的编程练习,照顾使用和保健的测试。软件的画出流程图见图3。格式化的模块的程序执行的用户界面。另一个模块带出去太。

Multi-transouter Sys

The program developed on the B004 system was then mapped onto a multi-transputer system as shown in Figure 4.The system comprised a B008 transputer development board having four TRAM modules, and the same speech synthesis board as before.The TRAMS each have a T414 transputer. Associated with two of the TRAM modules is 1 MByte of RAM and with the other two 32K Bytes of RAM. One of the 1 MByte TRAMS is situated in slot 0 of the BOO8 and runs the TDS. The mapping of the program used an adaptation of the processor farm model [10.11]. Unlike the process farm model which by its nature implicity assumes that a linear arrangement of workers is the optimal topology for a farm of processors. the model employed allows the passing of data to the two workers in parallel.The transputer in Slot 0 (the monitor) performs the interface between the transputer network and the microcomputer host. It also obtains the character strings to be matched by filtering the input text and updating the buffer; for subsequent string recognition.This is carried out concurrently by the transputers situated in slot 1 (worker one) and slot 3 (worker two), matching the input string to the rules within the rule set. For speed of execution both modules have a copy of all the rule sets. Worker one attempts to match on consonant/vowel, voiced consonant and sibilant.

Multi-transouter系统

这个项目发展的B004系统映射到一个multi-transputer系统就如图4。 该系统由B008 transputer发展委员会有四个电车模块,和相同的语音合成板和先前一样。每个人都有T414的有轨电车transputer。 有轨电车的两个相关模块是一个内存和所有与其他两个32K字节的内存。1所有的电车位于槽0的BOO8说明和运行。程序利用的映射的处理器的一种适应农场模型[10.11]。不像过程本身所造成的农场模型implicity假定一个线性安排员工是最优拓扑结构的处理器的一个农场。该模型通过允许使用数据到两个工人相似。transputer槽内的0(显示器之间的界面履行transputer网络和微机的主人。并获得匹配字符输入文本过滤,更新缓冲带;之后的字符串的认可。这是同时进行的transputers位于职工1(槽)和槽3(工人一式两份,输入字符串匹配规则规则集内。执行速度两个模块准备一份所有的规则集。职工试图配上辅音/韵母以外,表示辅音和字根。

characters and a buffer update module keeps a buffer full for the subsequent recognition modules. The recognition modules are bared around an algorithm operating upon a letter within a character string [9]. Each letter character has associated with it a unique set of rules, where each of the rules within the set contains a character string to be matched. These character matches are carried out in parallel by the transputer system as are the matches on the suffix and prefix characters for vowels, consonants, voiced consonants, sibilants, influencing consonants, front vowels and assorted matches. The concurrently running processes sequentially scan the rule set until the match is found. This is implemented by having a jump vector associated with

each rule set pointing to the start location of that rule set. associated with each rule in the set is its particular allophone string representation. The final module in the program passes the allophone string along a link via a channel to the speech synthesis board.

人物和缓冲更新模块保持一个缓冲满了随后的识别模块。识别模块在显露在信算法操作在一个字符串[9]。每一个字母与其相关人物有一组独一无二的规则,即每个规则在包含一个字符串匹配。这些字符并行进行transputer系统的比赛的前缀和后缀的元音、辅音角色,表达了,影响sibilants辅音、元音和辅音、前组合的比赛。同时运行进程的规则集扫描顺序,直至比赛发现。这是执行相关的有跳向量

每个规则集指着开始位置的相关规则set.在每一个规则是其独特的音位变体字符串表示。在程序中最后一个单元通过音位变体一直忠于通过渠道联系的语音合成板。

Figure 3. Standard pronunciation algorithm

A timer channel was introduced to time the execution of the program with and without the actual outputing of the allophone string to the speech synthesis board. These timings were carried out for strings of various lengths and types of characters, so as to be able to obtain a quantitative assessment of available computing time between allophone outputs. This time wuld be

used for further processing or performing communication. On average it vms found the processing teak 254.9 timer clock periods per character.The time including allophone output averages out to 1767.3 timer clock periods per character.

图3。标准发音算法

一个定时器通道被介绍给时间程序的运行实际用与不用上实现给水三冲量串级音位变体的细线语音合成板。这些时间中搜索字符串进行了各种长度和类型的角色,希望能够获得定量评价之间可资利用的计算时间音位变体输出。这次贵方觉得是

用于进一步的处理或表演的沟通。一般而言白银加工柚木254.9发现每个角色定时器时钟周期。时间包括音位变体输出每人每个角色1767.3定时器时钟周期。

A timer channel was again introduced to obtain a quantitative assessment of the processing time. On average it was found that the processing took 199.78 timer clock periods.

DISCUSSION AND CONCLUSIONS

From the timings obtained for the single transputer system it can be seen that for 86% of the time, the transputer system is idle waiting for the speech synthesis section to finish outputing allophones. In the case of the multi-transputer system. this idle time is further increased,as expected to 89% by the inclusion of more transputers. It was found that the speech quality was good though rather robotic due to the fixed accessing time of the speech synthesis chip and its lack of inflection. From these results, it can be concluded that a transputer system is ideally suited to this type of application. Work continues in this area using a more sophisticated speech synthesis chip, a more efficient form of parallelsim and morphological rules to facilitate the inclusion of stress and inflection. Work is also underway in the related area of speech generation using transputers to model the vocal system to eradicate the robotic nature of the sound from current speech synthesisers.

REFERENCES

W.A. Ainsworth, "A system for converting English

text into speech", IEEE Trans.

Electroacoust.

1974

Vol. AU-21,

pp. 288-290.

Audio

June

M.D. Mcllroy, "Synthetic English speech by rule".

&I1 Tele. Lab.. Inc., Murray Hill, NJ, Mar. 1974

F.F. Lee. 'Reading

一个定时器渠道再一次传入获得的定量评价处理时间。平均的加工,它被发现了199.78定时器时钟的时期。

讨论和结论

从时间单一transputer系统获得可以看出,为86%的时候,transputer系统是懒惰的等待语音合成技术部门完成变体上实现给水三冲量串级。在这个例子中是multi-transputer体系这空闲时间,如预期的那样进一步增加89%包含更多的transputers。结果表明,语音质量很好但机器人由于而固定访问的语音合成芯片和其缺乏调。从以上结果,它可以推断,transputer系统最适合这种类型的应用程序。在这个地区工作不断的使用更复杂的语音合成芯片,一个更有效率的形式的parallelsim和形态规则包含方便的压力和调。工作也进行语音的生成的相关领域的使用transputers声带系统模型根治的机器人本性的声音synthesisers从目前的演讲。

参考

W.A. Ainsworth,”一个系统转换为英语

文本为语言”,国立台湾科技大学。

Electroacoust。

1974年

页。AU-21,

288-290页。。

音频

6月

医学Mcllroy,“合成英语演讲由规则”。

&I1远程。实验室。墨累山公司,台北,Mar. 1974年

F.F.·李。 课外阅读