第16卷 第11期 小型微型计算机系统 Vol. 16, No. 11
1995年11月 MINI-MICRO SYSTEMS Nov., 1995

一个实用的汉语文语转换系统

倪 宏
(中国科学院声学研究所，北京 100080)

摘 要 本文介绍了一个基于语音参数规则合成的汉语文语转换系统。本系统采用汉语音节和词汇作为合成单元，保留了音节构词时音节与音节之间以及音节内部的超音段信息，保证了合成语音的自然度；采用目前较成功的CELP语音编码方法对合成单元进行压缩，在压缩20多倍的情况下仍能保证合成语音的高清晰度。作者在构建系统时对系统软件的完善考虑以及对用户编程接口的设计，使得该系统成为一个有广泛用途的实用的汉语文语转换系统。

关键词 语音合成，文语转换系统，规则，CELP

1 引 言

文语转换系统是人机智能语音接口的重要组成部分。它最终实现的功能是，对于任意的文本输入，计算机能够给出自然流畅的汉语语音输出。国内的一些科研单位一直在开展这方面的研究工作，并取得了很多成果[1]。尽管目前由于存在着理论和方法上的难点，合成语音的自然度距离句子一级的自然流畅尚有距离，但受到电子信息和排版等行业的实用需求所促进，国内一些从事汉语语音合成研究工作的单位纷纷将其研究成果实用化，有些产品已经初步商品化。

文语转换系统是语音合成技术在广度和深度上的延伸和拓展，它所涉及的不仅包括语音学、语言学、心理学等传统研究学科方向，还包括计算机科学、电子学和通信等高科技研究领域。文语转换系统的总体结构按处理的次序可分为两个层次，第一层次是文字处理部分，对文本进行分析，包括分词、意群修正，完成字位至音位的转换，同时得到句子的韵律模式；第二层次是语音处理(语音合成)部分，它根据文字处理的结果，将音位单元重构，并加入语音学的韵律规则，输出连续的话音。语音合成从处理方法上可以归纳为两大类，一类是时域的波形编辑合成，另一类是频域的参数规则合成。波形编辑合成采用一些时域的处理技术如基音同步重叠相加(PSOLA)技术，直接对音位单元的波形进行压缩和恢复，得到合成话音，其特点是运算量小，建立实时合成系统时不需专用的DSP部件，易于实现，产生的合成话音的音质普遍较高，但语音库所占的存储量较大。频域的参数规则合成基于语音产生模型，提取合成单元的语音参数，如LPC，共振峰等参数打包建库，在合成语音时根据语音韵律规则

收稿日期：1994-11-30。本文由中国科学院"八五"重大科技攻关项目资助。倪 宏，助理研究员，主要从事语音处理、图像处理、DSP软件技术的研究。

修改上述参数，通过合成滤波器完成语音的重构，发出合成话音。参数合成的特点在于以语音产生模型为基础，参数的物理意义清楚，规则的加入和参数的修改更为直观和方便，存储量小。它所产生的合成话音的音质普遍低于波形编辑合成，但近几年随着人们对语音处理研究工作的不断深入，语音编码技术取得了重大的突破和长足的进展，基于码激励线性预测的中低速率语音编码成果已经广泛实用化，它的语音音质已经得到了人们的首肯。这也为参数文语转换系统推向社会起到了促进作用。

我们在汉语参数规则合成工作的基础上[2]，研制了一个实用的汉语文语转换系统 CELP－TTS。

2　系统设计

本系统在 286/386/486 PC 机或兼容机上实现，内插一块 DSP 专用处理板作为合成器的硬件，系统软件包含语音库、系统基本分词词库、用户词库、特征词库及规则库、合成器等，分属于文本分析和合成语音两个模块。CELP－TTS 的系统结构如图 1 所示。

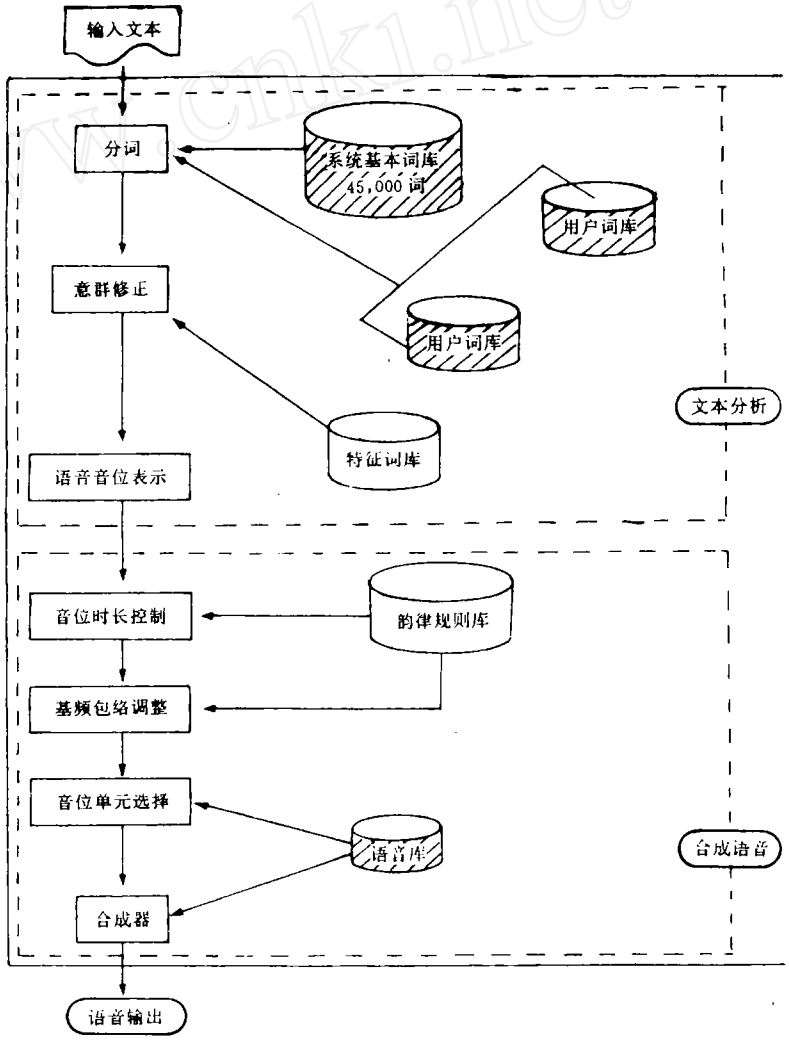

图 1　CELP－TTS 的系统结构

2.1　文本处理模块

对于微机中的汉字文本文件，该模块根据系统分词词库和用户词库对其进行分词处理，根据特征词库对分词结果进行意群修正(包括给出多音字、词的初选和提示，语句的间歇和意群等信息)，得到语音的音位表示和韵律信息。

2.1.1　分词

文本分词是汉语文语转换系统中的一个重要组成部分，由于汉字书面表示中，只是一个个由标点符号分开的汉字字符串，无法体现“词”的概念，所以对于文语转换系统来说，首先必需将这些汉字字符串“切割”成由“分词单元”组成的链，得到与语音库中合成单元的映射关系，以控制语音合成的内容。“分词

单元"储存在系统词库和用户词库中。系统词库包含了现代汉语词汇45,000个，词长2—4。(用户词库是为用户扩展而提供的，其大小和词长由用户确定，但处理方法与系统词库相同)。采用"自右向左最大匹配算法"进行自动分词，最大匹配长度为词库中最长词的长度。若没有相匹配的词，则认为是单字的组合。该算法的分词准确率在99%以上。

2.1.2 意群修正

口语与书面语不同，尽管书面语中的每个句子由词和单字组合而成，但是人们在口语中朗读句子时，并非单纯地以词或单字为单元发音，而是按照句子的长短和内容对词和单字进行组合，形成句子的节奏，这一方面受人们常说的"语感"的制约，另一方面也是人们短时记忆的惯例。在汉语中语音的节奏与意群是吻合的，文章，包括诗歌和散文，每个句子可以分割成若干个意群，意群与意群之间有明显的停顿。意群形成了语音的韵律。分词结果加入意群修正能够更好地反映句子级语音的韵律包络，提高句子的合成语音自然度。对意群的分割目前尚没有一个绝对的标准，在实际工作中，我们采用以下几种规约的组合来切割意群：

(1) 助词如"的""得""吗""哪""啦""了"等，如果不在词中，与前一词或字在同一意群，并为当前意群的结束；

(2) 副词、形容词如"很""太""较""在"等，如果不在词中，与后一词或字在同一意群，并为当前意群的开始；

(3) 四音节语段给人以稳定的感觉，所以被广泛采用；

(4) 使用过程中对特征词库不断修正，使意群修正符合专业的特点。

例如本文的题目分词结果为：

一个 实用 的 汉语 文语 转换 系统

意群修正后结果为：

一个 实用的 汉语文语 转换系统

后者更符合口语的习惯。

分词和意群修正的好坏直接影响着输出语音的自然度[3]。从某种意义上说，对于系统自动处理的结果，人工修正是重要的。人工修正的结果保存在特征词库中。同时，对于书面语中出现的多音字和多音词，系统能够根据特征词库按出现频率给出初选和提示；能够将句子中的数学符号和阿拉伯数字转换成相应的发音单元等，从而给出句子正确发音和节奏的控制信息。

2.2 语音处理模块

2.2.1 语音库

结合汉语本身特点，本系统选用音节和高频词汇作为基本合成单元。汉语共有有调音节1281个，对应于国标二级字库6768个汉字，音节作为汉语文语转换系统的基本合成单元是必需的。为了克服音节与音节之间在构成词汇语音时的不自然，我们在选用音节的同时，还选取汉语中常用的词汇作为基本合成单元，保留了词汇中音节构词时音节与音节之间以及音节内部的超音段信息，提高了输出语音的自然度。我们根据现代汉语词汇的使用频度，选取双字词16800个，三字词2100个，四字词1500个，以及单音节1281个，共两万余个合成单元[4]。这些单元由播音员发音、8kHz14位A/D采样得到原始语音数据。

采用CELP编码方法对原始语音数据压缩建库。码激励线性预测(CELP)是介于波形编码和参数编码之间的一种语音编码方法，近年来广泛用于移动通信等语音数字编码领域，并形成国际标准。它可以在每秒4800比特的压缩速率上得到较好的音质[5]，我们认为这种合成

话音用在文语转换系统中是可行的，本系统使用该方法对词汇语料库进行压缩，构成本系统的基本语音库，压缩率约为二十五倍，保留的参数有表示声道模型的 PARCOR 系数、声激励源参数和基频参数等。

2.2.2　规则库

汉语是声调语言，节奏感强，在语流中，人的听觉对句子一级音调包络的敏感程度远远大于对词一级音调包络的敏感。规则库总结了人们的发音习惯和发音规则，给出对合成单元语音参数加以修正的规则和方法，用于提高篇章语句的自然度。

在语音的线性模型中，线性预测系数(或 PARCOR 系数)是发声时声道的物理表示，激励源参数描述了对声道的刺激。发声时声道的形状及其变化速率有限，表现在线性预测系数的相对稳定，人们听觉上比较敏感的音调反映在浊音段激励源参数的准周期性上。通过调整激励源参数的幅度改变合成语音的轻重；通过增加或减少浊音段准周期性的激励源参数的周期个数改变合成语音的长短；通过调整浊音段激励源参数准周期的长度改变合成语音的高低。上述理论为规则库中规则模式的实现提供了操作方法和理论根据。规则库主要包括：

(1) 时长规则。调整句间间隔、意群间隔、意群内各合成单元的时长及其之间的间隔。

(2) 变调规则。汉语韵律的魅力来自于变调。变调包括上上相连、阴阴相连、阳阳相连、去去相连、“一、七、八、不”变调等。

(3) 轻声规则。轻声在语流中轻声的变化对于人的听觉和语感来说比较敏感，并且有辨义的功能。

2.2.3　合成器

合成器采用以 TMS320C3x 或 TMS320C25 芯片为主芯片的数字信号处理板，用以完成语音实时重构，语音输出等功能，该芯片带有自己的汇编语言，语音实时重构及输出是由它的汇编程序完成的。

3　系统实现及用户接口

本系统是一个 DOS 环境下的独立的系统软件，用 MSC5.1 语言编写。系统的主界面是

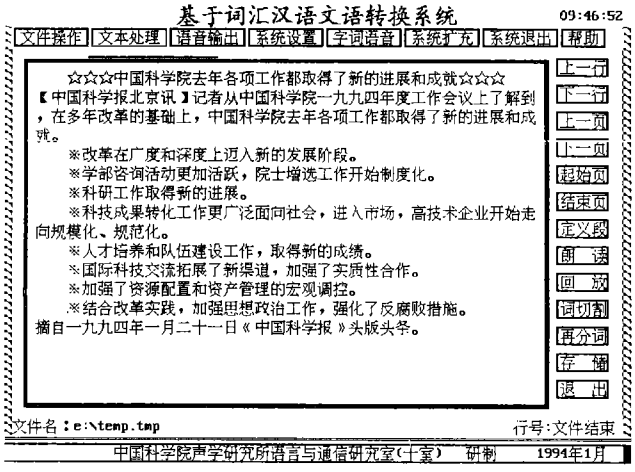

图 2　CELP－TTS 的人机界面

一个彩色图形汉字下拉式主菜单，主菜单包含有文件操作、文本处理、语音输出、系统扩充

等。用户在该系统上加载汉字文本文件，经自动分词和意群修正送合成器，即可实时发出语音。系统的人机界面如图2所示。

除完成上述设计思想之外，系统扩充模块为用户设计了应用编程接口(API)，以增强本系统的实用性：

(1) 用户可根据使用需要增加语音库。不同领域的用户其常用词汇不尽相同，用户可将自己常用的词汇甚至短语(长度不限)的发音录制下来，8kHz14位采样A/D转换为磁盘数据，使用系统提供的"增加词库"功能把语音压缩至语音库中，并把相应的汉字加进用户分词词库中，随后的文本中出现这些新的词或短语时，它们将会作为合成单元对待。

(2) 用户根据自己的习惯修正或添加特征词库，作为以后意群修正的依据。

(3) 用户在自己的应用程序中增加语音功能。本系统带有附加库，库中提供了语音功能所需的过程和函数，用户在自己的程序中调用这些函数，并与附加库连接即可实现语音输出。附加库中的函数包括：文件和汉字句子的分词处理，文件句子和词汇的语音输出等。

(4) 利用本系统可以构成有限语句的语音输出系统。

4 结束语

本系统采用汉语词汇作为合成单元，保留了词汇中音节与音节之间的超音段信息，使输出语音有较高的自然度和清晰度；友好的软件界面和强大的系统扩充功能，本系统作为一个实用的汉语文语转换系统有广阔的推广前景。随着规则研究工作的不断深入，逐渐扩充语音规则，将能够得到更为流畅自然的语音输出。

参 考 文 献

〔1〕 张家录. 汉语文语转换系统的语音规则和声学参数. 声学学报，第15卷，第2期，pp.113－119，1990

〔2〕 Hong NI et al., Chinese Speech Rule－Synthesis System, Proc. WPRAC Ⅳ, pp.189－195, 1991

〔3〕 刘源等. 现代汉语常用词词典，宇航出版社，1990

〔4〕 李彤等. 汉语规则合成的字音流自动转换，第六届全国语音图像通讯信号处理学术论文集，ppb7.99－b7.102，四川，1993

〔5〕 Schroeder and Atal, Code－Excited Linear Prediction(CELP): High－Quality Speech at very Low Bit Rates, Proc. ICASSP, 1985

USEFUL CHINESE TEXT－TO－SPEECH SYSTEM

Ni Hong

(*Institute of Acoustics, Academia Sinica, Beijing* 100080)

Abstract Chinese text－to－speech system based on speech parameteric synthesis by rules is presented in this paper. This system utilizes Chinese syllables and words which is constituted of syllables as synthetic units, reserves the supra－segmental features between syllables and within syllables, this method guaranteed the synthetic speech naturalness. Using presently successfully CELP speech coding method to compress the synthetic unit, under the ratio of 20, can still have high articulation synthetic speech. Thorough consideration to the software development and application program interface design while constructing the system will make it a widely used practical Chinese text－to－speech system.

Key words Speech synthesis, Text－to－speech, Rules, CELP