

Design and Application of Speech Synthesis Technology Used in Alarm Systems

Zhenjiang Zhao¹, Weikang Zhao^{1,2}, Dantong Na¹, Yuqiao Liu¹, Haishan Zhao¹

¹Changchun Institute of Engineering Technology, Changchun, China

²Changchun University of Science and Technology, Changchun, China

Email: na198073@yahoo.cn, s060400628@yahoo.cn

Abstract: This paper presents a new application of the alarm systems designed by speech synthesis technology. Currently available alarm systems (such as in elevators, hospitals and other places) are a single form, which do not have the humanized effect, showing no specific information. In order to solve this problem, the system uses the speech synthesis technology, installs the TTS (Text to Speech) speech synthesis chip in the alarm systems which are controlled by using single-chip micro-computer. It can change the simple digital signals, text messages and alarm information into voice information that can be more easily converted into acceptable ones, and the direct use of voice manner to the rescue staff released by a variety of distress signals, so that rescue personnel provide a more efficient and safer handling of emergency situations.

Keywords: alarm systems; speech synthesis; text to speech

语音合成技术在呼救、报警系统中的设计应用

赵振江¹, 赵维康^{1,2}, 那丹彤¹, 刘玉桥¹, 赵海山¹

¹长春工程技术学院, 长春, 中国, 130117

²长春理工大学, 长春, 中国, 130117

Email: na198073@yahoo.cn, s060400628@yahoo.cn

摘 要: 本文提出了一种新的应用语音合成技术设计制作呼救、报警系统的方法。目前现有的呼救、报警系统(如在电梯、医院等场所)形式单一、不具备人性化效果、展现不出具体状况。为解决这一问题, 利用语音合成技术, 将 TTS (Text to Speech) 语音合成芯片加装在呼救、报警系统上, 采用单片微型计算机进行控制, 将简单的数字信号、文本信息和单音报警信息转换为更易于让人接受的话音信息, 直接用语音方式向救援人员发布各种求救信息, 使救援人员能够更快捷、更安全的处理突发情况。

关键词: 呼救、报警系统; 语音合成技术; 文语转换 (TTS)

1 引言

语音合成技术是将计算机自己产生的或从外部输入的文字、符号、数字等信息转换成自然流畅的语音信号输出, 这种技术又称文语转换技术 (Text to Speech, 简称 TTS)。如果能将语音信号合成技术应用在呼救、报警系统中, 把电梯报警系统的故障情况或医院病房求救系统的病人需求的文本信息用人类的语音表示出来, 由于需求具体化了, 可以更加灵活地处理各种突发状况, 有利于节省救援时间, 使求助人员处于更加安全的境地; 有利于医院或应用电梯的部门提供更多人性化的服务; 有助于新技术的融合和发展。本文

提出了一种新的应用语音合成技术的呼救、报警系统的实现方法。本文在广泛汲取现代语音信号处理和汉语韵律知识的基础上, 对语音合成方法及其实现进行了深入的研究, 结合 TD-PSOLA 算法的优势, 提出这种新的呼救、报警系统。

2 语音合成技术

语音技术, 能够为信息时代人与人之间的沟通及人与机器之间的沟通提供一种简便、方便和有效的手段。由人工制作出的语音称为语音合成 (Speech Synthesis)。

2.1 文语转换系统

语音合成技术是将计算机自己产生的或外部

输入的文字信息,按语音处理规则转换成自然流畅的语音信号输出,使人们通过“听”就可以明白信息的内容。这种将文字转换成语音的技术也称文语转换技术,简称TTS(Text to Speech)技术。文语转换(TTS)是指把文本文件通过一定的硬软件转换后由计算机或其他语音系统输出语音的过程。

文语转换系统实际上可以看作是一个人工智能系统。一个成功的文语转换系统输出的语音应当音质清晰,自然流畅。因此,文语转换(TTS)系统的核心应包含文本分析、韵律控制、语音合成三个模块^[1]。图1显示了一个完整的文语转换系统示意图。

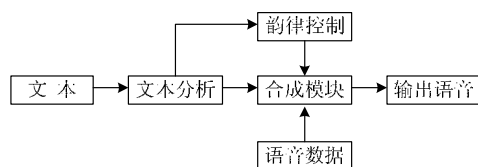


图1. TTS 文语转换系统基本框架

2.2 文本分析模块

文语转换系统首先处理的是文字,也就是它要说的内容。文本分析的主要功能是使计算机能够识别文字,并根据文本的上下文关系在一定程度上对文本进行理解,从而确定要发什么音、怎么发音,并将发音的方式告诉计算机,另外还要让计算机知道文本中哪些是词,哪些是短语、句子,发音时到哪儿应该停顿,停顿多久,等等。文本分析模块将输入的文字转换成计算机能够处理的内部参数,送给韵律生成模块进一步处理并生成相应信息。

2.3 韵律控制模块

在自然语流中,人们使用语调、节奏和重音等方式来表达说话者的意向和情感,这些特征就称为韵律(Prosody)。它是语言交际的必要手段,能帮助听者更好地理解话语所携带的信息。语调、节奏和重音这些韵律特征是通过超音段特征——音高、音长、音强及频率分布的变化表现出来的,因此,这些超音段特征的修改成为韵律控制的基础。韵律控制模块的任务就是要计算言语中的韵律特征,建立有效的韵律模型,以产生具体的韵

律参数,为下一步的语音合成提供依据。

2.4 语音合成模块

在得到语音的内容和韵律参数后,TTS系统的最终目的是利用声学模块合成出语音来。应用于呼救、报警系统的语音合成方法主要是波形拼接合成法。

波形拼接合成法以语句、短语、词或音节为合成单位,经录音、编码压缩后形成语音库。这种语音合成技术用原始语音波形替代参数,而且这些语音波形取自自然语音的词或句子,它隐含了声调、重音等细微特性,合成出的语音清晰自然,质量高。

3 语音合成算法

3.1 基音同步叠加算法 PSOLA

PSOLA 算法的原理是:将原始语音信号与一系列基音同步窗相乘得到一系列短时分析信号:将短时分析信号修正后得到短时合成信号,根据原始语音波形信号和目标波形的基音曲线和音长,确定二者之间的基音周期映射,从而确定所需的短时基音序列;将合成的短时基音序列与目标基音周期同步排列,重叠相加得到基音波形,此时合成的语音波形就是所期望的基音周期曲线和音长^[2]。

3.2 TD-PSOLA 语音合成算法

我们主要研究时域基音同步叠加(TD-PSOLA),TD-PSOLA 算法只在时域对波形进行处理,不对信号作频域上的调整。音高(基频)的修改是通过改变基频之间的时间间隔得到的,音长的修正通过重复或者省略一些语音片断来完成。

基音同步叠加(PSOLA)算法能直接处理语音波形的基频和时长,它的实质是将短时傅里叶变换方法中声源滤波器分解和声源谱修改这两个步骤合为一步来完成。

信号 $x(n)$ 的短时傅立叶变换为:

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)\omega(n-m)e^{-j\omega n}, n \in Z \quad (1)$$

由于语音信号是一个短时平稳信号,因此在时域每隔若干个(例如 R 个)样本取一个频谱函数就可以重构信号 $x(n)$,即可令:

$$Y_r(e^{j\omega}) = X_n(e^{j\omega})_{n=rR, r, r, n \in Z} \quad (2)$$

其傅立叶逆变换为:

$$y_r(m) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y_r(e^{j\omega}) e^{j\omega m} d\omega, m \in Z \quad (3)$$

然后就可以通过叠加 $y_r(m)$ 得到原信号, 即:

$$y_r(m) = \sum_{r=-\infty}^{\infty} y_r(m) \quad (4)$$

时域基音同步叠加(TD-PSOLA)技术作为基音同步叠加技术的一种, 通过以下步骤实现语音合成:

(1) 基音同步分析

同步标记是与合成单元浊音段的基音保持同步的一系列位置点, 同步标记可以准确的反映基音周期的起始位置。以语音合成单元的基音同步标记为中心, 选取适当长度的时窗对合成单元作加窗处理, 获得一组短时信号 $X_m(n)$:

$$X_m(n) = h_m(t_m - n)X(n) \quad (5)$$

式中, t_m 为基音标注点, $h_m(n)$ 一般取 Hamming 窗, 窗长大于原始信号的一个基音周期, 因此窗间有重叠。系统中窗长取为原始信号的周期的 2 倍。

(2) 基音同步修改

同步修改在时长规则的指导下, 调整基音同步标记, 产生新的基因同步标记。具体地说, 就是通过对合成单元同步标记的插入、删除来改变合成语音的时长; 通过对合成单元标记间隔的增加、减少来改变合成语音的基频等。这些短时合成信号序列在修改时与一套新的合成信号基音同步标记。在 TD-PSOLA 方法中, 短时合成信号是由信号直接拷贝而来。若短时分析信号为 $X(t_a(s), n)$, 短时合成信号为 $X(t_s(s), n)$, 则有:

$$X(t_s(s), n) = X(t_a(s), n) \quad (6)$$

式中 $t_a(s)$ 为分析基音标记, $t_s(s)$ 为合成基音标记。

(3) 基音同步合成

基音同步合成是利用短时合成信号进行叠加合成, 如果合成信号仅仅在时长上有变化, 则增加或减少相应的短时合成信号; 如果是在基频上有变换则首先将合成信号变换成符合要求的短时合成信号再进行合成。基音同步叠加合成的方法有很多, 我们采用原始信号谱与合成信号谱差异最小的最小平方叠加合成法 (Least-Square

Overlap-Added Scheme)。最终合成信号为:

$$\bar{X}(n) = \sum_q a_q \bar{X}_q(n) \bar{h}_q(t_q - n) / \sum_q \bar{h}_q^2(t_q - n) \quad (7)$$

式中, 分母是时变单元化因子, 是窗之间时变叠加的能量补偿, 这里 $\bar{h}_q(n)$ 为合成窗序列, a_q 为相加归一化因子, 是为补偿高音修改时的能量损失而设的。(7)式可以简化为:

$$\bar{X}(n) = \sum_q a_q \bar{X}_q / \sum_q \bar{h}_q(t_q - n) \quad (8)$$

式中的分母作为时变的单位化因子, 在窄带条件下接近于常数, 在宽带条件下, 当合成窗长为合成基音周期的 2 倍时该因子亦为常数。此时, 若设 $a_q=1$, 则有

$$\bar{X}(n) = \sum_q \bar{X}_q(n) \quad (9)$$

利用公式(7)和(8), 可以通过对原始语音的基音同步标记间的相对距离的修改, 即可完成对合成语音的基音进行提升或降低, 还可以通过对音节的基音同步标记的插入和删除来实现对合成语音音长的改变。同样可以调整能量因子 a_q 来调整语流中不同位置的合成语音的输出能量。

3.3 基音同步标记

基音同步标记对语音合成单元作基音同步标记处理是 TD-PSOLA 算法的关键。短时信号的截取、时变时窗长度的选择以及合成语音基频的变化、时长的增减等, 都是依据基音同步标记进行的。每个基音标注文件与音库中对应的音节波形编码文件一一对应, 其中内容是音节波形的基音周期的个数和每个基音周期的起始点的位谱序列。采用短时 AMDF 方法为音节库中的每个汉字标注基音同步标志。

3.4 韵律调整的语音合成主要步骤

基于 TD-PSOLA 算法韵律调整的语音合成主要步骤:

(1) 根据韵律规则生成的韵律控制符得到所需要的基频时长因子;

(2) 对音库中的基音同步标志作变换, 包括: 插入和删除基音同步标记、提高或降低标记间隔;

(3) 用汉明窗与原始语音波形相乘(窗长取基音周期两倍)得到一系列的短时信号, 然后按照变换后的基音同步标志进行叠加, 计算出需要播放

的波形。

4 系统设计

本文提出了一种基于语音合成技术的呼救、报警系统的设计方法。系统采用 TTS 语音合成芯片，将求助人员发送给救援人员的信息、文本转换为具有高自然度的语音，直接用语音方式向救援人员发布各种信息。该系统具有概念先进，适用范围广泛，方便实用的特点。

4.1 系统结构

图 2 为系统结构框图。其中 PC 机位于该系统的控制中心，是呼救、报警系统的管理主机，呼救、报警系统管理软件中整合了信息发布平台；F 为楼层控制机，由单片机进行控制；R 为护士/电梯管理人员监控室分机，用来接收 PC 机发布的信息数据；A 为病房/电梯。系统采用联网式呼救、报警系统的布线网络，并兼容其全部功能。系统中的信息传输通道采用了 CAN 总线，提高了数据通信的实时性、可靠性和安全性，较好地解决了所发布的呼救信息与报警信息的碰撞问题。系统工作时，由求助人员在病房或电梯键入需要发布的信息文本，或用指示器选择计算机中预置的信息，然后点击发送按钮，将信息发出。根据实际情况的需要，可选择对具体某一监控室进行信息发布。所发布的信息经 CAN 总线接口模块进入信息传输通道。楼层控制机接收总线上的信息并进行判断，对楼层控制机地址范围内的信息文本进行转发，否则不进行处理。护士/电梯管理人员监控室分机接收到楼层控制机转发来的信息文本数据后，在单片机的控制下进行信息存储，并通过语音合成芯片将其转换为语音信号，经放大后驱动扬声器发出声音。

4.2 软硬件设计

系统硬件利用联网式呼救、报警系统的布线网络，通信干线采用 CAN 总线。CAN 总线接口模块用于两种不同通信标准之间的转换。楼层控制机的硬件结构与病房/电梯硬件结构基本相同。护士/电梯管理人员监控室分机加装了语音合成芯片，采用单片微型计算机进行控制，用于将病房/电梯发送来的字符信息转换为语音信息。该部分的硬件结构如图 3 所示。

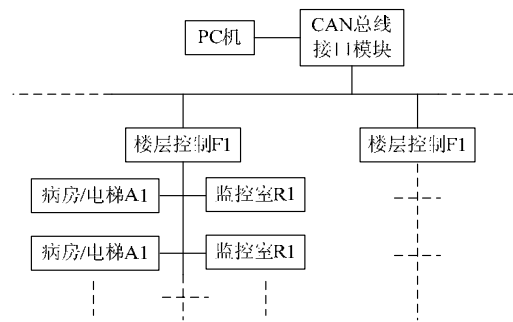


图 2. 系统结构

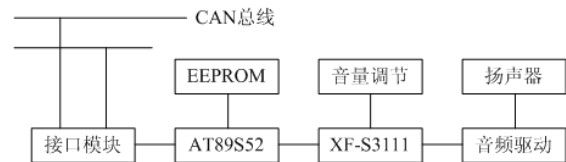


图 3. 监控室分机硬件结构

图 3 中，AT89S52 为 8 位单片计算机，其内部带有 8K 字节的 Flash ROM，用于存放运算程序。XF2S3111 为科大讯飞公司生产的采用 InterSound3.0 语音合成系统的 TTS 语音合成芯片，可在单片机的控制下完成文本到语音的转换。它通过异步串口接收待合成的文本，直接合成为语音输出，具有清晰、流畅、自然的中文语音合成效果。EEPROM 的作用是对信息文本数据进行存储，从而使系统信息可以即时发布。转换后的语音信号经音频放大器放大后驱动扬声器发出声音。语音合成器是一独立的功能块。当主处理机用 RS-232 口告知语音合成器有语音输出要求时，语音合成器的 CPU 从相应的串口数据寄存器读取信息，并进行判断，按要求产生输出语句。语音合成器的主要任务是实时地响应主处理机送来的命令字，并把固化在语音参数存储器的语音参数按命令字进行解码回放，针对语音合成器需要的音质要求高，词汇量小等特点，语音合成拟采用波形编码技术，这种方法相对简单、易行，合成的语音质量好^[3]。

系统软件分为上位机的信息发布及管理软件和监控室分机的信息接收及文语转换软件。上位机软件采用 VC++6.0 编写^[4]，主要包括信息发布界面、信息数据库、信息管理、与楼层控制机的通信等部分；下位机软件采用单片机汇编语言编写，主要完成信息的接收、存储、与楼层控制机和 TTS 芯片的通信及文本信息输出等功能。监控室分机软件的主

程序框图见图 4。

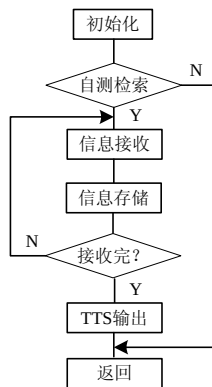


图 4. 监控室分机主程序流程

语音合成软件由自测试模块、控制模块及合成模块组成。自测试模块：完成语音合成模块的自测试工作，包括指令系统，程序 EPROM，语音参数 EEPROM 及整个放音系统的测试工作。控制模块：控制模块实现对语音合成器的控制管理工作。包括语音合成器的初始化工作，与主处理机的通信管理工作及输出语音信息的检索、判断工作以及词组的

编排、组织工作。

5. 结语

将语音合成技术应用于呼救、报警系统，使该系统具有概念先进，适用范围广泛，方便实用的特点，可以更加灵活地处理各种突发状况，有利于节省救援的时间，使求助人员处于更加安全的境地；有利于医院或应用电梯的部门提供更多人性化的服务；有助于新技术的融合和发展。系统采用联网式综合布线网络，并兼容其全部功能，应用方式灵活，具有较好的应用前景。

References (参考文献)

- [1] Chen Guilin, Yue Dongjian, Zu Yiqiang, YuZhenli. An Embedded English Synthesis Approach based on Speech Concatenation and Smoothing. ISCSLP2004, Hong Kong, 2004.12.
- [2] XF2S3111 Speech Synthesis Chip Data Sheet Rev1.0. 2003.09.
- [3] Li Minghua. Application of AT89C2051 Used in Speech Synthesis. Computer and Modernization, 2005(5):P78-80.
- [4] Fang Jianchun. Integrated System of Speech synthesis technology and MCU [M], Beijing University of Aeronautics and Astronautics Press, 1993.