基于 DTW 算法的语音识别与仿真

张　静，齐国红

郑州西亚斯学院，河南 新郑 451100

摘　要　在孤立词语音识别中，动态时间规整 DTW 算法是一种应用较为广泛的算法之一，有着较强的科学性，在立足于当前 DTW 语音识别算法应用的实际情况下，简略阐述了该课题的研究背景，并从预处理和特征参数提取以及 DTW 算法两方面着手对基于 DTW 算法的语音识别系统实现进行了探究，以此为基础展开了相应的仿真和分析，旨在为相关研究人员提供参考。

关键词　DTW 算法；语音识别；系统实现；仿真分析

中图分类号　TP29　　**DOI**　10.19769/j.zdhy.2021.09.021

0 引言

从不同的角度出发，语音识别有着多样化的分类，根据说话人讲话的方式有所不同，可以将其划分成连续语音识别、连接词识别以及孤立词识别三种类型，对于孤立词识别来说，其每一次只需要一个短语或者是词汇，并将其看作是词汇表中的相应的词条。在当前社会发展以及现实生活中，孤立词语音识别的应用极为广泛，最关键的则在于自动控制方面，具体包括智能玩具、仪器设备操作以及家用电器控制等等，在人们无法实现手动控制的情况下，语音控制便能够发挥出其应有的作用，基于此，有必要对其展开更加深层次的探究。

1 研究背景

在进行语音识别的过程中需要先对语音的特征进行选定，将其看成是识别参数的模板，接下来便可以采用一个能够对参考模板以及未知模板似然度进行衡量的测度函数，然后便可以对专家知识以及最佳测度准则进行选定，使其作为识别决策最后实现对于识别候选者的判定。但由于说话人不能够两次对同一个词产生完全相同的发音，其差异性既体现在频谱的偏移以及音强的大小等方面上，还体现在发音时其音节长度的不同方面，与此同时，说话人两次发音的音节并会呈现出较为突出的线性对应关系。基于线性时间规整技术假设说话人的说话速度主要是从不同说话单元实际的发音长度出发进行等比例分布的。此外，其同所说的声音之间呈现出相互独立的特点，由此可见，测试模板同参考模板之间所具有的畸变度量主要指基于平面上的矩形对角线展开计算的，通常情况下来说会采用欧几里得距离对 T 中的第 i 帧特征以及 R 中的第 i 帧特征之间的畸度变量进行表示。

$$i_n = \frac{N}{M} i_m$$

结合实际情况来看，上文所述针对说话速度差异所进行的限制难以同具体的语音发音情况相适应，所以应当尽量采用一种能够更适应现实条件的语音时间规整方法。对于语音识别来说，动态时间规整作为重要的算法之一有着较强的应用价值。当构建和应用小词汇表孤立词识别系统时，其所具有的识别率与相应的指标基本上等同于采用 HMM 算法所得出的结果，但从实际情况来看，DTW 算法本身有着较强的高效性和简便性，所以往往能够应用在一些特定的场合当中。然而 DTW 动态搜索方法对于矩阵的存储能力有着较高的要求，若是对其展开直接的计算工作势必会涉及大量空间的占用并产生较高的计算量。基于上述所存在的各方面问题，本文针对 DTW 语音识别算法展开更加深层次的探讨，并基于此对其采取了相应的优化改进措施，通过对于宽度限制动态经搜索方法的应用，实现了语音识别速度的进一步提升，并基于各种仿真适应分析比较了本文方法以及传统方法之间的差异，对本文所采用的方法的科学性和有效性进行了验证[1]。

2 基于 DTW 算法的语音识别系统实现及仿真分析

2.1 系统实现

2.1.1 预处理和特征参数提取

预加重主要指的是通过对于一阶高通数字滤波器的应用来达到提升高频特性的效果，在加窗分帧方面，窗函数主要是对汉明窗进行应用，为了能够更加方便地实现对于参数的计算，其帧长和帧移分别取 256 和 80。端点检测的应用实现了在以往所应用的基于过零率和短时能量双门检测的基础上的进一步优化调整，与此同时还额外增加了 2 个门限值，分别为最短语音时间门限值以及最大静音门限值。其中静音门限的设置能够有效起到防止漏检的作用，通常情况下来说会取 3－5 帧之间，而语音时间门限的设置则可以达到将突发性噪声滤除的效果，一般会取 9－10 帧，具体还要从实际情况出发对门限值进行设定。本系统主要是将梅尔倒谱系数以及 MFCC 差分倒谱系数共同作为其相应的特征参数，因为人的耳朵在动态特性方面有着更加敏感的特点，所以不仅要使用 MFCC 参数同时还需要能够对语音动态变化进行充分

收稿日期：2021-09-02

作者简介：张静（1983—），女，汉族，河南新郑人，讲师，硕士研究生，主要从事自动化研究；齐国红（1987—），女，汉族，河南淮阳人，讲师，硕士研究生，主要从事数字图像处理。

反映的差分倒谱参数，MFCC 的参数主要选 12 阶，而差分倒谱则去其一阶差分，这样将其合在一起便是 24 阶。

2.1.2 DTW 算法

结合实际情况来看，语音识别本身便是模板匹配的过程，针对误差更小的参考模板进行搜寻以及匹配，并在此基础上得到相应的识别结构，测试模板主要指的是其要进行识别的一个输入词条语音，而参考模板则主要是在模板库中所包含的每一个训练好的词条。若是采用 C 和 F 分别对测试模板以及参考模板进行表示，便可以用二者之间的实际距离 D 对两者之间所具有的相似度进行度量，若是其相似度更高其距离便更小，若想真正实现对于这一失真距离的科学计算，需要从 C 和 F 中每一个对应帧之间所具有的距离着手展开针对性的计算。假设在 F 中 m 为任选的帧号，m＝1～M，n 则是 C 中任选的帧号，其取值范围在 1～N 之间，与之相对应的两针特征矢量距离用 d 来表示，与此同时，N≠M。DTW 算法的应用本质上有两种功能，一方面是针对 C 和 F 所对应的各帧之间的距离进行合理计算，另一方面便是从帧匹配距离出发对最佳路径进行精确搜寻。结合上述研究能够明确，若是采用传统 DTW 算法，势必会产生过大的运算量，往往要通过多次的运算才能够获得一个相应的时间弯折函数。结合实际情况进行分析能够明确，在对 DTW 算法进行实际应用的过程中通常会在其中加入相应的搜索限制条件，而具体的搜索范围主要是处在相应宽度范围之内的，结合相关研究可以知道，在经过相应区域限制之后的 DTW 的动态搜索宽度大多是会局限在对角线附近的带状区域当中。

综合分析上文所述的相关内容之后，笔者决定对搜索宽度限制的存储空间分配方法进行应用，假设其搜索的宽度是 Width，那么其仅仅需要分配 3×Width 的存储空间，基于这个范围针对动态规划路径进行全面搜索，对累计匹配距离展开计算工作，此举能够尽量实现对于存储空间的减少，并达到减小计算量的效果。这样一来便可以得出相应的局部约束路径，假设两个需要进行比较的模板所采用的帧移、窗函数、帧长、特征矢量相同，其中纵轴主要指的是帧数较少的，而帧数较多的则是横轴[2]。Width 具体是 2×(N—M)，从矩形平面的左下角着手进行匹配，直到右上角完成，在完成每一次的横轴模板矢量的比较工作之后，实验人员都需要针对 d_1、d_2、D 中的数据展开相应的更新升级工作，直到纵轴模板最后一帧 M 便可以停止，由此可以判定待测模板同参考模板之间的距离是 Min。

2.2 仿真分析

本文基于搜索宽度限制 DTW 方法，开展了相应的仿真实验，并希望通过分析验证其科学性和有效性。本文所做的实验使用的数据基本上来源于计算机的声卡录音，其采样频率为 8 000 Hz，语音数据都是在不同的时间段所进行录音的。在样本集一中的数据主要包含着播放、停止、打开、关闭、前进以及倒退等，这些都属于语音控制媒体播放系统的相关之指令词，其总共包括 20 组样本，都是女生发音，这也代表着每一个指令次都具有 20 个具有差异性的发音，样本总共有 120 个。而在样本集 2 中的数据则都属于单个字，分别是放、停、开、关、前、后。这些数据也都是女生发音，共 15 组，实验数据一共包括 90 个语音。每开展一个试验都将第一组的发音看成是训练模板，并以其余的几个样本为待测数据。在语音特征参数方面主要使用的是依靠听觉系统的 MFCC，每帧语音都会 24 个特征参数进行提取。

本实验主要是针对搜索宽度限制 DTW 算法以及区域限制 DTW 算法的实际性能展开对比分析，经过分析之后能够发现，二者在识别速度方面并没有产生过大的差异，但结合实验结果可以得出结论，相对于整体路径约束 DTW 算法来说，搜索宽度限制 DTW 算法有着更高的识别率。本实验的最终结果判定，搜索宽度限制 DTW 算法的应用可以在原有的基础上实现对于存储空间的减小，并进一步促进识别准确率以及识别速度的提高。实验研究结果表明，搜索宽度限制 DTW 算法有着更快的识别速度，之所以会如此主要是因为其识别算法有着相对较小的计算量，常规的 DTW 算法往往涉及 M×N 次帧匹配，但搜索宽度限制的 DTW 算法仅仅涉及 M×2(N—M)，假设 M 和 N 分别为 40 和 50，那么普通 DTW 算法往往需要在整个的平面上展开 2 000 次的帧匹配计算工作，但若是能够应用搜索宽度限制的 DTW 算法便可以在原有的基础上将帧匹配计算次数缩减到 800 次[3]。

如果其横轴以及纵轴语音帧数之间不存在过大的差异，那么还可以进一步减少识别算法原有的计算量，与此同时，搜索宽度限制的 DTW 算法不会涉及过大存储空间的应用，这便在一定程度上缓解了常规 DTW 算法所面临的硬件系统难以实现的现象，一般来说，普通的 DTW 算法有着 2×M×N 的空间需求，以满足对于帧匹配距离矩阵以及累计距离矩阵的存储需求。但应用搜索宽度限制的 DTW 算法仅涉及对于 3×2×(N—M)存储空间的应用，若是 M 和 N 分别为 40 和 50，便能够明确搜索宽度限制的 DTW 算法仅仅需要常规算法存储空间的 1%。在对二者之间性能进行分析和讨论的过程可以进一步验证搜索宽度限制的 DTW 算法的有效和科学性，能够在孤立词语音识别算法中得到高效应用。

3 结语

综上所述，针对 DTW 算法展开改进和优化调整工作能够进一步减少其在存储空间方面的限制，对于语音识别过程的优化有着积极的促进作用。因此，相关研究人员需要加强对于其的重视，进而更好地实现对于 DTW 算法的应用，助力系统运行效果的提升。

参考文献

[1] 王语涵，闫子薇，武天琦，等.基于 DTW 算法对婴幼儿语音的分析[J].齐齐哈尔大学学报：自然科学版，2019，35(4)：69-73.

[2] 王素宁，朱俊杰，李志勇，等.基于 DTW 算法的电力调度语音识别研究和应用[J].电力与能源，2021，42(1)：35-38，64.

[3] 黄奕婷，于宝雲，高丽萍，等.基于 LPC 模型的 DTW 语音转换系统设计[J].信息通信，2021，34(1)：71-74.