利用动态时间规整方法对说话人识别的实验研究

常国岑　刘国军

（信息工程学院）

动态时间规整方法是一种以动态规整为基础的模式匹配方法，在模式识别特别是语音识别领域中有着广泛的应用。本文介绍将动态时间规整方法应用于说话人识别的方案，实验研究的结果表明，动态时间规整方法应用于说话人识别是行之有效的，而且实验的错误率可以降低到静态时间规整说话人识别实验的错误率的一半。

一、引　　言

语音信号中包含的信息可以划分为韵质信息（共性信息）和声质信息（个性信息）两大部分[1]。语音识别的任务是利用语音中的共性信息识别说话的内容，而识别说话人的目的则是利用语音中的个性信息识别说话的人。

模式匹配是统计模式识别的一个重要环节，在进行说话人识别或语音识别的时候，即便对于同一实验文本，不同说话人的发音或同一说话人不同次的发音节奏的快慢和持续时间的长短方面也会发生很大的变化，从而很难将待识模式和参考模式进行匹配。语音模式的时间规整即是为解决语音模式的匹配问题而采取的一种措施。

二、系统概述

我们以IBM PC/XT微机为基础构成了一个说话人识别的实验系统。输入语音经前置放大、抗混滤波、模数转换之后成为数字信号，再经端点检测、特征提取、时间规整获得参考模式或待识模式，最后经模式匹配和识别判决输出识别结果。

在该识别系统上同时进行了两组实验：

抗混滤波器的截止频率为4.5kHz，模数转换器的采样频率为10kHz。实验文本的语音时长约300ms，在进行线性预测分析时帧长取30ms，帧与帧之间重迭10 ms。

三、结果分析

说话人识别实验一和实验二的结果分别如表1和表2所示。

可以看出，动态时间规整说话人识别的错误率可以比静态时间规整说话人识别的错误率降低近一半，说明将动态时间规整方法应用于说话人识别系统是行之有效的。

四、结　　语

本文简要地阐述了将动态时间规整模式匹配方法应用于说话人识别的系统构成和实验结果，证实了将动态时间规整模式匹配方法应用于说话人识别的可行性与有效性。

说话人识别在司法（如案件审理）、军事（如情报侦察）、保卫（如身份证明）、金融（如电话服务）等方面有着广阔的应用前景。我们的实验研究只是初步的探讨，有待于继续改进和不断完善。

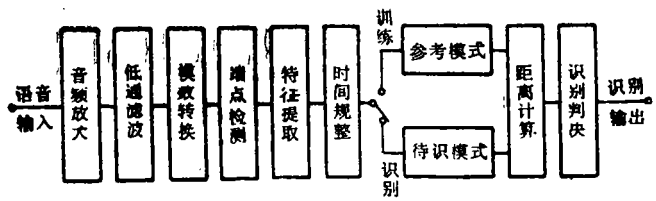


图1 说话人识别实验的系统构成

表1 说话人识别实验1的实验结果

参考语音 / 待识语音		静态规整 S_1	S_2	S_3	S_4	S_5	参考语音 / 待识语音		静态规整 S_1	S_2	S_3	S_4	S_5
男	S_1	19	1	0	0	0	男	S_1	18	1	0	1	0
男	S_2	1	18	1	0	0	男	S_2	0	19	1	0	0
男	S_3	1	2	17	0	0	男	S_3	1	0	19	0	0
女	S_4	0	0	0	20	0	女	S_4	0	0	0	20	0
女	S_5	0	1	0	0	19	女	S_5	0	0	0	0	20

录音条件	实验文本	说话人数	发音次数	特征参数	规整方法	总错误率
普通实验室	字音"二"重复五遍	男三人 女二人	各20次 共100次	平均基音频率 线性预测系数	静态时间规整	7.00%
					动态时间规整	4.00%

表2 说话人识别实验2的实验结果

参考语音 / 待识语音		动态规整 S_1	S_2	S_3	S_4	S_5	S_6	参考语音 / 待识语音		动态规整 S_1	S_2	S_3	S_4	S_5	S_6
男	S_1	15	0	0	0	0	1	男	S_1	16	0	0	0	0	0
男	S_2	2	12	2	0	0	0	男	S_2	0	16	0	0	0	0
男	S_3	0	0	14	0	2	0	男	S_3	0	1	14	0	1	0
男	S_4	0	0	1	14	0	1	男	S_4	0	0	0	14	1	1
男	S_5	0	0	0	0	16	0	男	S_5	0	0	0	0	16	0
男	S_6	0	0	0	1	0	15	男	S_6	0	0	2	0	0	14

录音条件	实验文本	说话人数	发音次数	特征参数	规整方法	总错误率
普通实验室	字音"二"重复五遍	男六人	各16次 共96次	平均基音频率 线性预测倒谱	静态时间规整	10.42%
					动态时间规整	6.25%