基于DF-MAP说话人模型训练

王　军

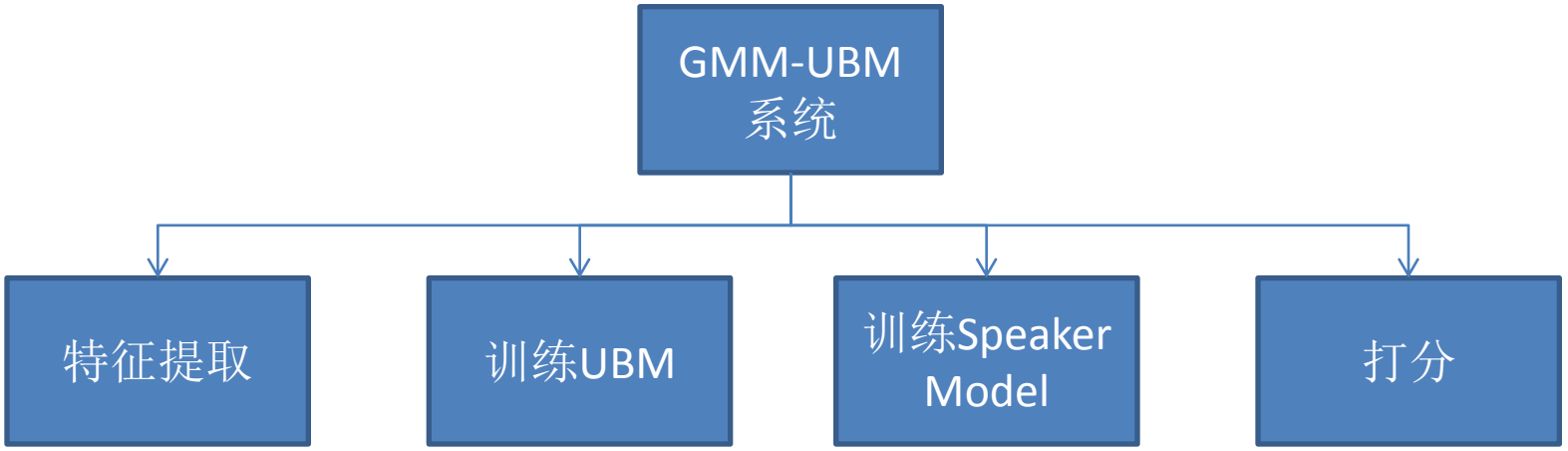
CSLT，RIIT，THU

2013-01-06

目 录

1、基于GMM-UBM说话人识别

- 自NIST1999说话人确认评测以来，GMM-UBM由于其出色性能成为说话人确认的最主要方法。[1]



1、基于GMM-UBM说话人识别

特征提取：

分帧，预加重，提取特征，去除静音。

采用MFCC特征。DFT，三角带通滤波与离散谱卷积求对数能量，DCT。[1]

训练Speaker Model：

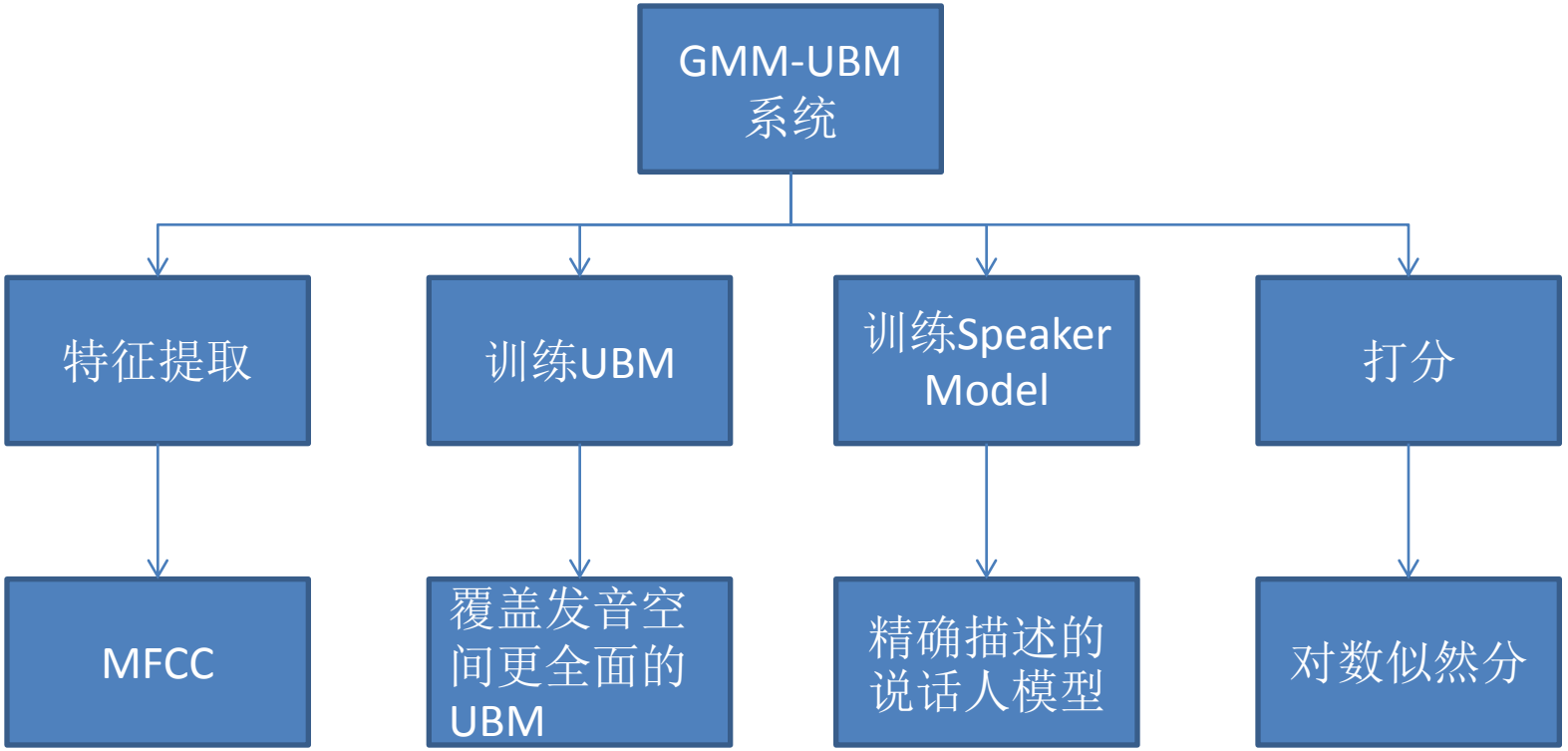
在GMM-UBM系统中，说话人模型通过贝叶斯自适应方法修改UBM中某些参数得到。[1][3]

训练UBM：

UBM是一个说话人无关、高阶的GMM。通常由数百人、男女均衡、数小时语音训练而成，用于表示说话人无关的特征分布。[1]

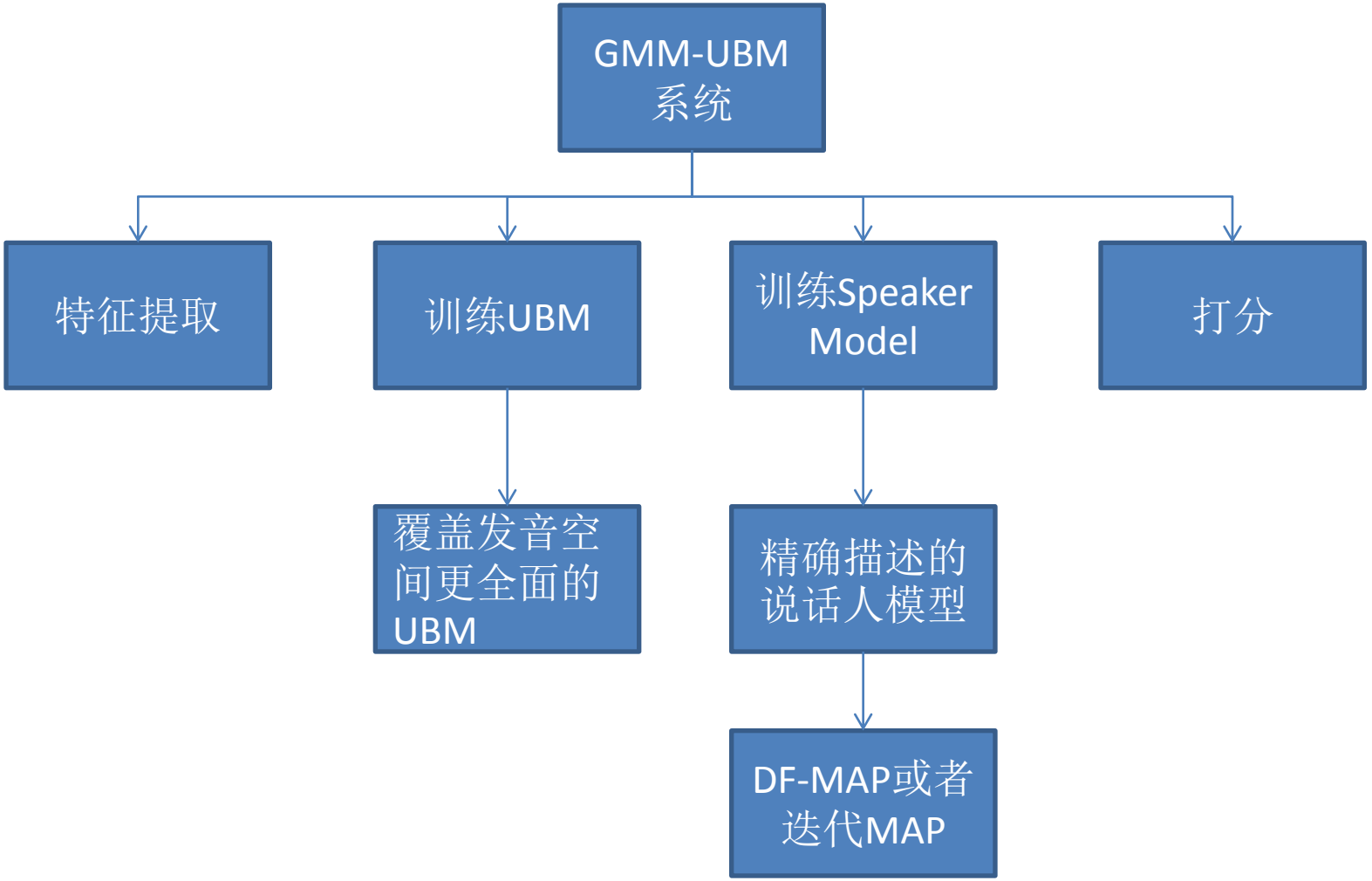
1、基于GMM-UBM说话人识别

➢GMM-UBM系统



2、基于DF-MAP的说话人模型训练方法

➢GMM-UBM系统性能



2、基于DF-MAP的说话人模型训练方法

说话人模型的训练： [1][3]

在GMM-UBM系统中，说话人模型通过贝叶斯自适应方法修改UBM中某些参数得到。

第一步是期望过程，计算训练数据在UBM各单高斯分布上的统计参数；

第二步，用新的统计参数与UBM的参数加权得到说话人模型的参数。

$$\tilde{\mu}_i = \alpha_i E_i(x) + (1 - \alpha_i)\mu_i$$

$$\alpha_i = \frac{n_i}{n_i + r}$$

2、基于DF-MAP的说话人模型训练方法

DF-MAP说话人模型的训练： [1][3]

说话人训练语音覆盖到的发音情况，可以用自己的语音建模；未覆盖到的发音情况，可以用说话人无关的特征分布近似，从而减少训练语音和测试语音不同的影响。[1]

（1）训练语音较短时，一次MAP可能导致说话人语音训练不充分。

（2）训练语音中说话人语音分布不均匀，导致对于某些发音训练充分，而另一些发音训练不充分。

3、实验

实验环境配置

- 时变声纹数据库[2]（方法针对建模，不考虑时变因素）

- VPR5.0

- UBM
  - UBM1：CCB_Adapt_110720.ubm
  - UBM2：UBM_WFCC_Reading_32.ubm
  - UBM3：男45人女38人,共83人的语音训练
  - UBM4：利用训练语音对UBM3自适应

3、实验

表1 不同MAP factor对说话人识别结果的影响

	CCB_Adapt_110720	UBM_WFCC_Reading_32
Baseline MAP=16	15.271610%	6.107486%
MAP=0	**----**	7.116808%
MAP=1	**13.433333%**	6.553955%
MAP=2	13.641808%	5.940254%
MAP=4	14.052542%	**5.707910%**
MAP=8	14.633333%	**5.705508%**

3、实验

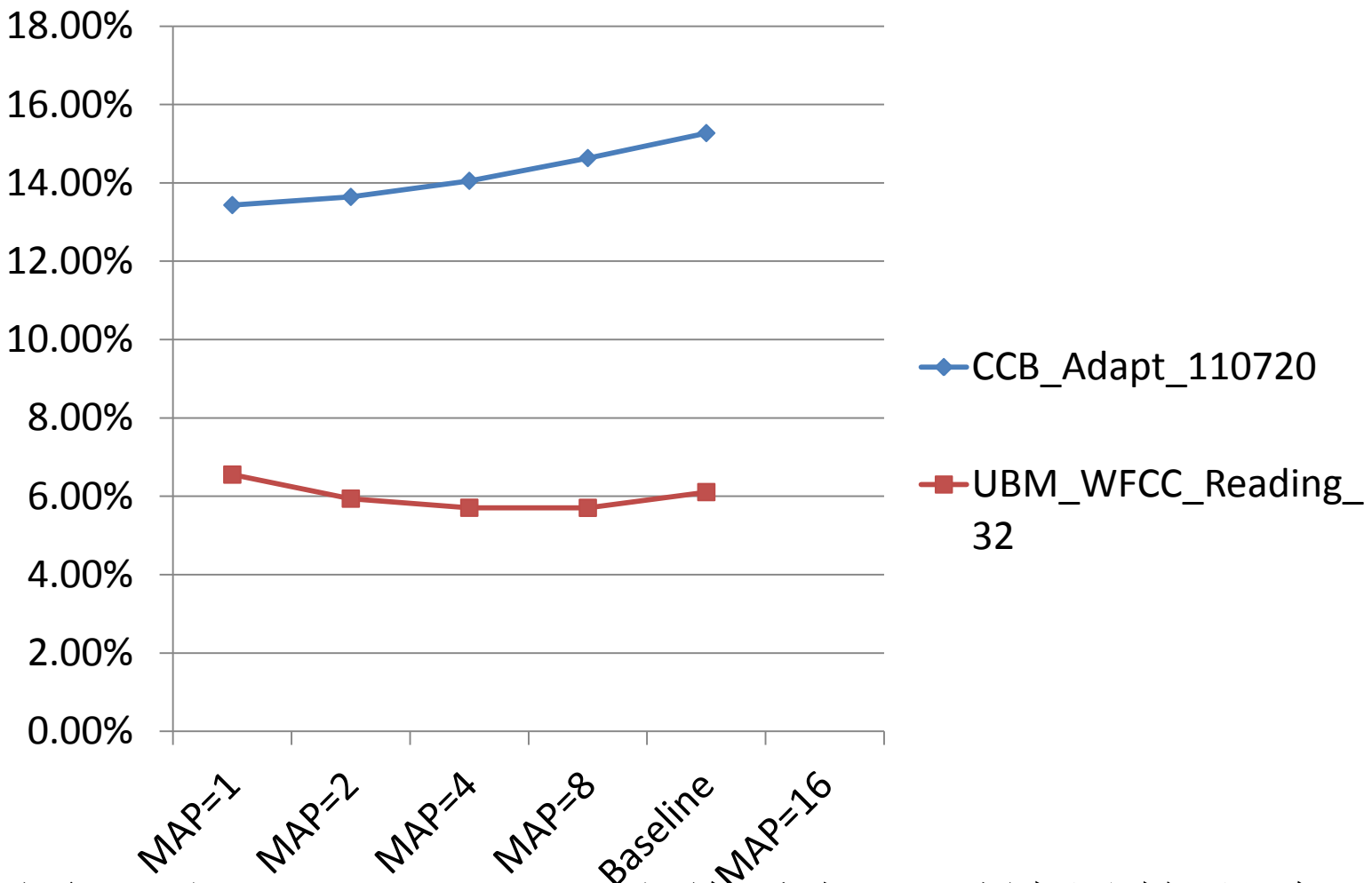


图1 不同MAP factor对说话人识别结果的影响

3、实验

表2 MAP迭代算法对说话人识别结果影响

	CCB_Adapt_110720/ utterance	UBM_WFCC_Reading_32/ utterance
baseline	15.271610%	6.107486%
迭代1	14.120198%	**5.734887%**
迭代2	13.284463%	5.867090%
迭代3	12.817090%	6.334322%
迭代4	12.605650%	7.502542%
迭代5	12.587429%	10.266667%
迭代6	**12.556497%**	14.002684%
迭代7	12.600000%	17.966667%
迭代8	12.687429%	22.433333%
迭代9	12.816808%	25.916667%
迭代10	12.933333%	29.153672%

3、实验

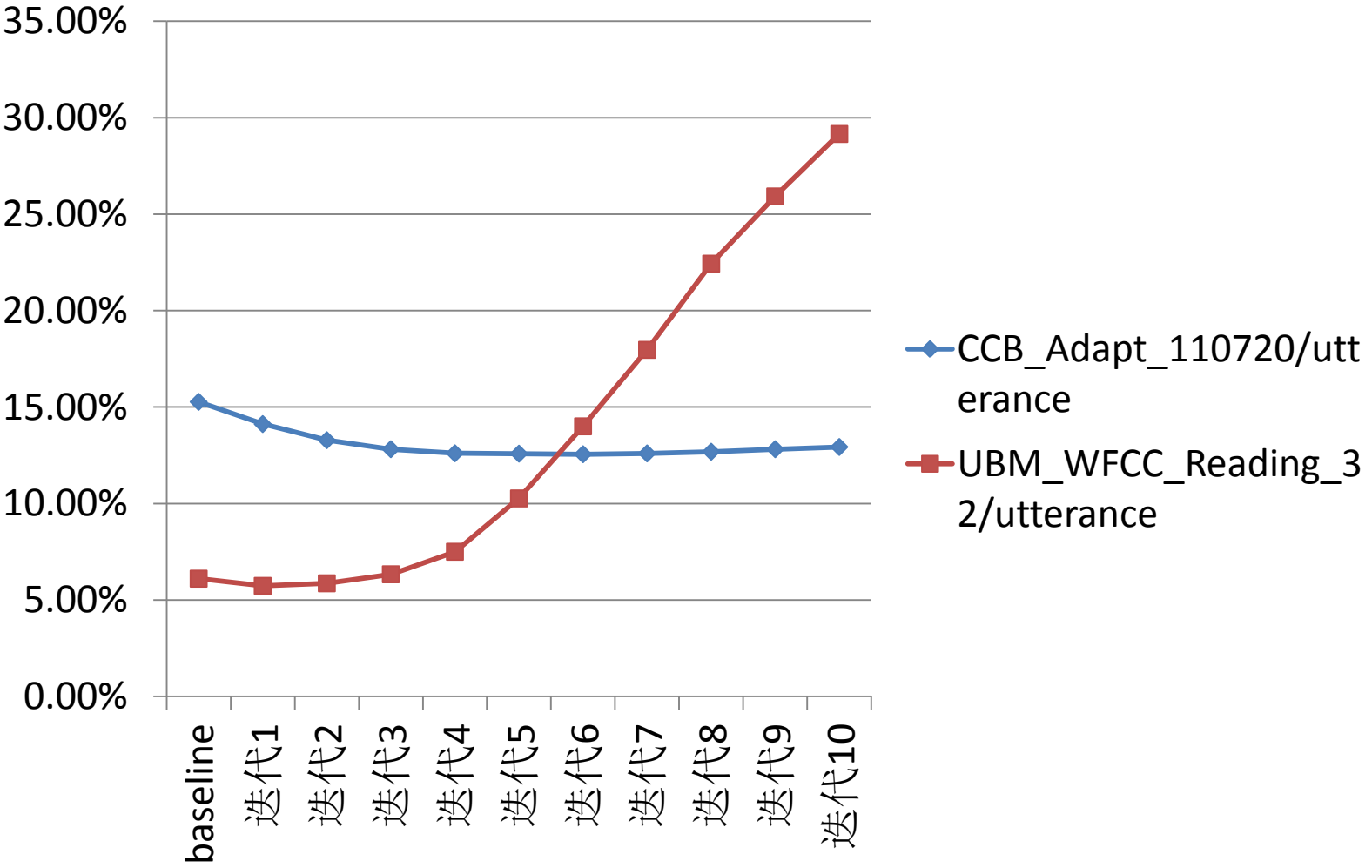


图2 MAP迭代算法对说话人识别结果影响

3、实验

表3 DF-MAP算法对说话人识别结果影响a

	数据1	**数据3**	**数据4**	**数据5**	**数据6**	**数据7**
baseline	11.969492%	12.258399%	14.125108%	13.373025%	13.432591%	14.570763%
DF-MAP	10.200000%	10.352573%	12.344298%	11.804367%	11.766844%	12.800000%
Relative improvement	14.78%	15.54%	12.61%	11.73%	12.40%	12.15%

3、实验

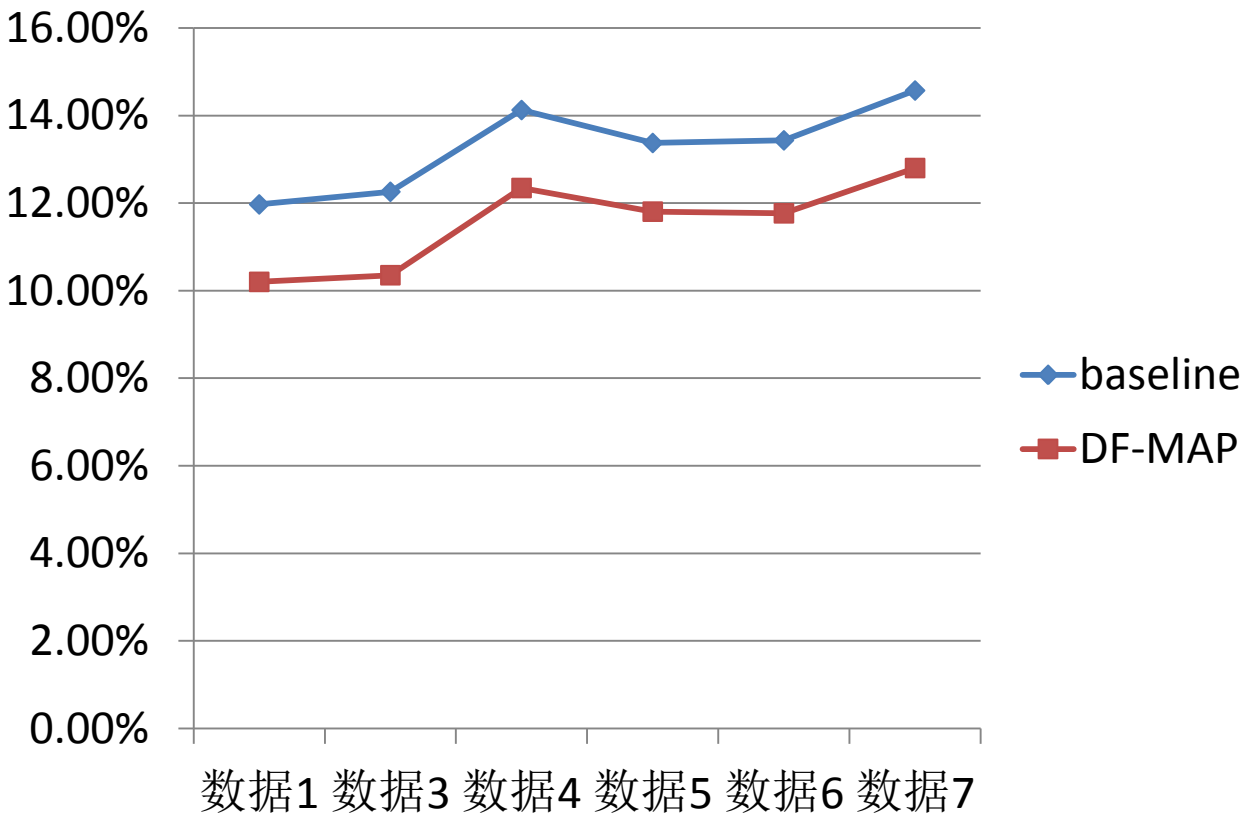


图3 DF-MAP迭代算法对说话人识别结果影响a

3、实验

表4 DF-MAP算法对说话人识别结果影响b

	数据1	数据2	数据4	数据5	数据6	数据7
baseline	9.273023%	6.107486%	11.173944%	9.884229%	8.801317%	11.740537%
DF-MAP	8.583333%	5.588277%	10.762712%	9.677966%	8.528315%	11.100000%
Relative improvement	7.44%	8.5%	3.68%	2.09%	3.1%	5.46%

3、实验

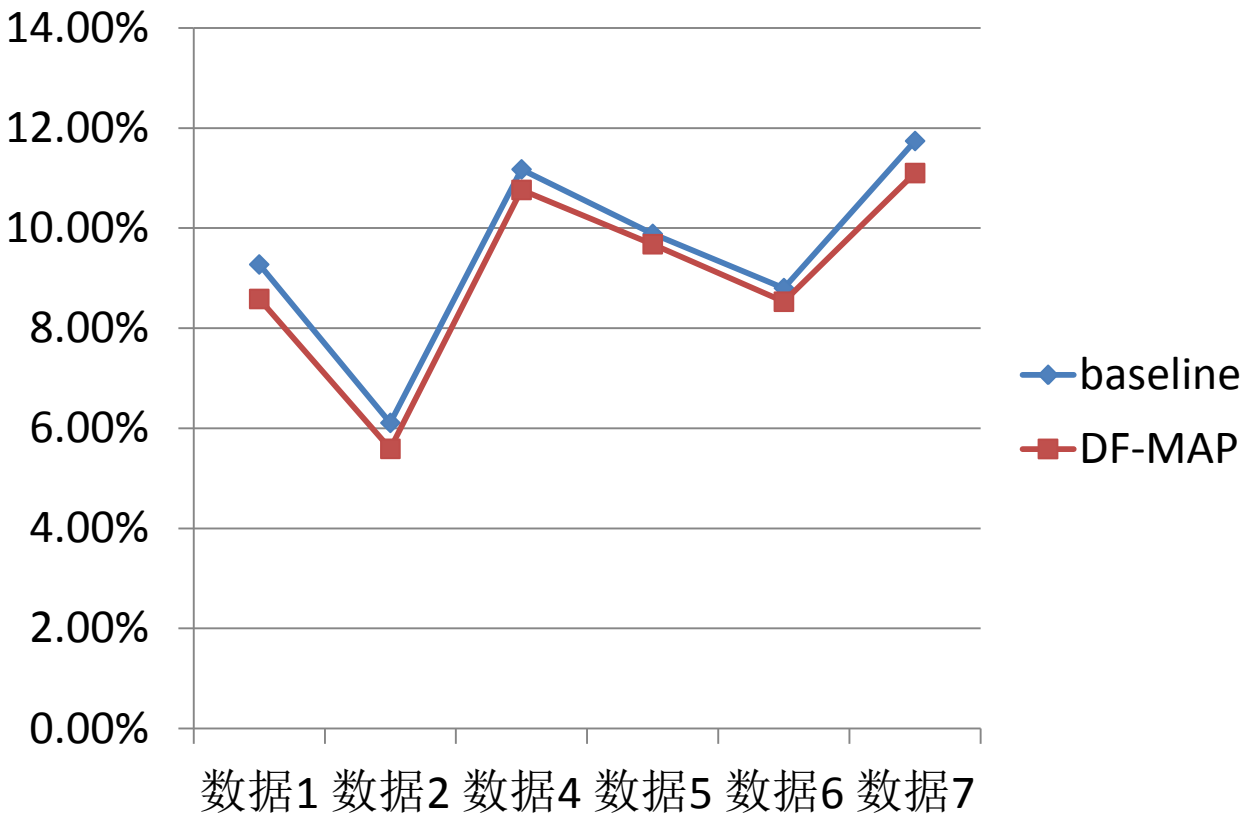


图4 DF-MAP对说话人识别结果影响b

4、结论

1、通过DF-MAP和MAP迭代方法可以对训练语音进行均匀、充分的训练，从而得到更精确的说话人模型，提高说话人识别性能。

2、最佳MAP迭代次数与UBM相关。不同的UBM会导致不同收敛速度，可依据开发集确定迭代次数。

3、最佳DF-MAP因子和训练语音相关。不同的训练语音分布导致不同的DF-MAP因子。可依据开发集确定DF-MAP因子。

。

参考文献

[1]熊振宇,大规模、开集、文本无关说话人辨认研究,:[博士学位论文].北京:清华大学计算机科学与技术系,2005.
[2]Linlin Wang and Thomas Fang Zheng, “Creation of Time-Varying Voiceprint Database,” Technical Session-6 (Oral), Oriental-COCOSDA, Nov. 24-25, 2010, Kathmandu, Nepal.
[3]Douglas Reynolds, “Gaussian Mixture Models,” Encyclopedia of Biometric Recognition, 2008.

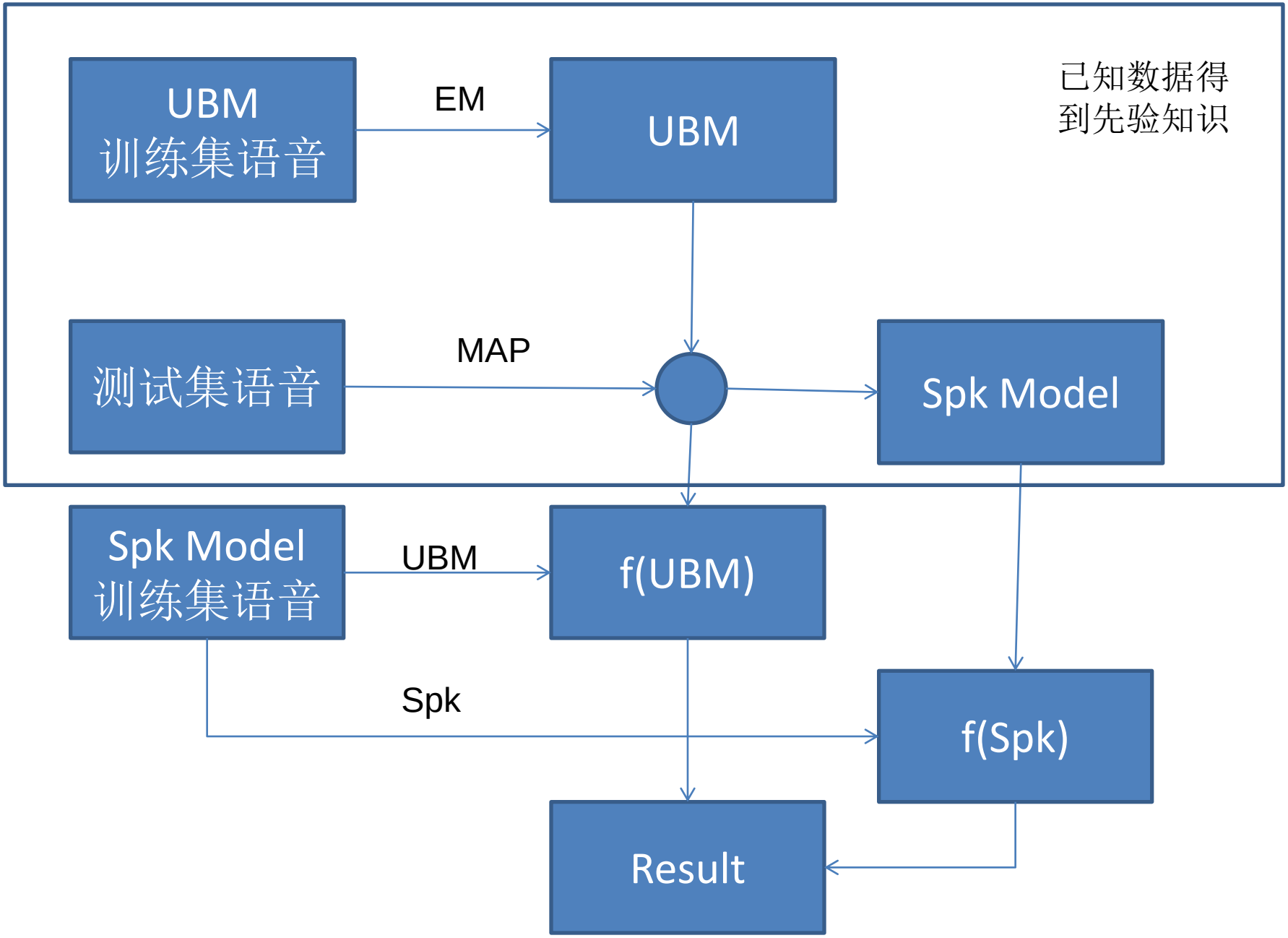
UBM训练方法

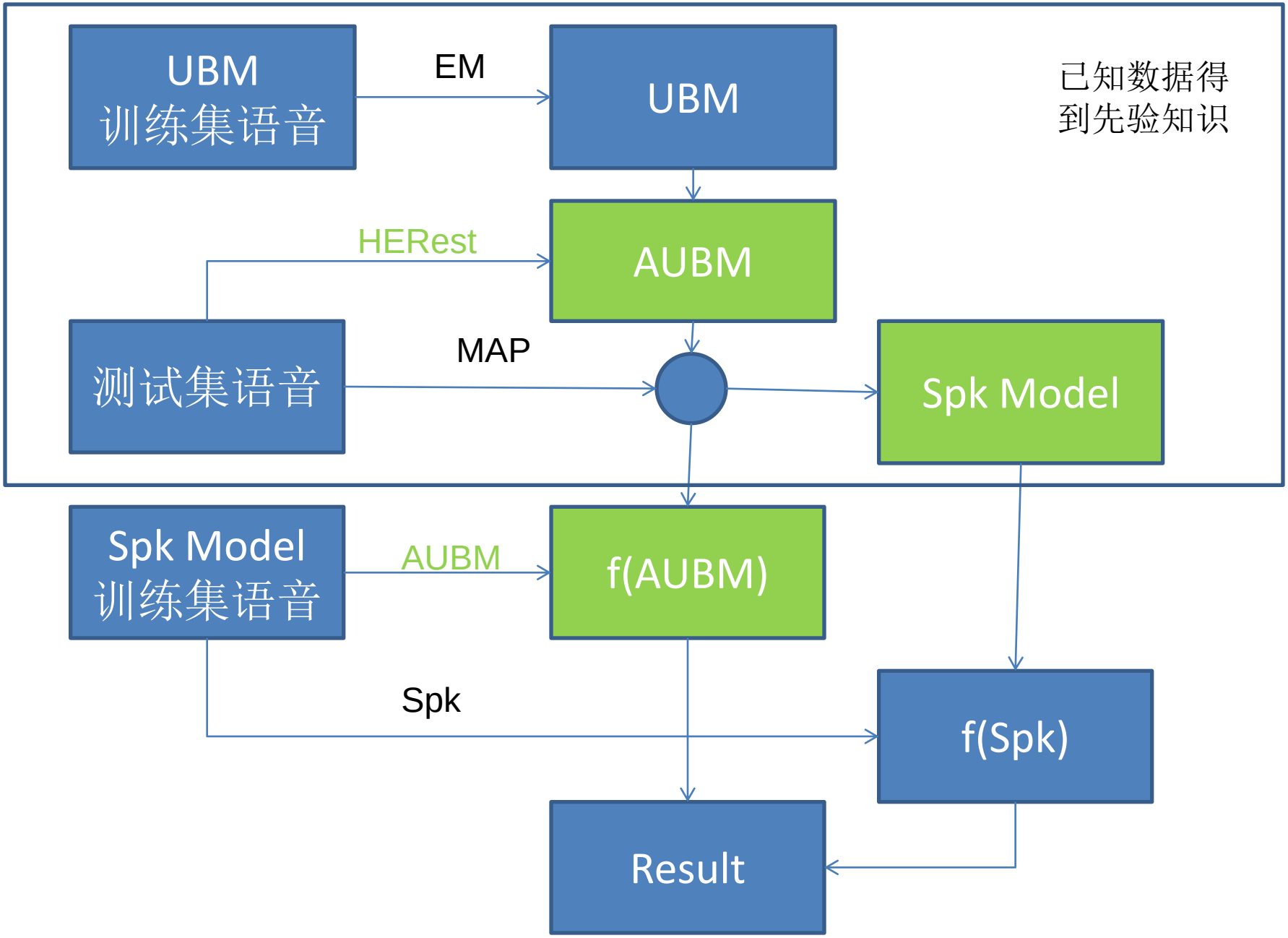
UBM是一个说话人无关、高阶的GMM。通常由数百人、男女均衡、数小时语音训练而成，用于表示说话人无关的特征分布。

（1） UBM训练需要较多人数、同信道语音，数据难找。

（2）UBM训练费时较长。

利用训练语音自适应可以得到一个在训练集上局部最优的UBM，且耗时大幅降低，性能大幅提高。


UBM
训练集语音
EM
UBM
已知数据得
到先验知识
测试集语音
MAP
Spk Model
Spk Model
训练集语音
UBM
f(UBM)
Spk
f(Spk)
Result


UBM
训练集语音
EM
UBM
已知数据得
到先验知识
HERest
AUBM
测试集语音
MAP
Spk Model
Spk Model
训练集语音
AUBM
f(AUBM)
Spk
f(Spk)
Result