一种稳健的基于 Visemic LDA 的口形动态特征及听视觉语音识别

谢 磊 付中华 蒋冬梅 赵荣椿
Werner Verhelst* Hichem Sahli* Jan Conlenis*
（西北工业大学 计算机学院 西安 710072）
* (Dept of ETRO, Free University Brussels, Pleinlaan 2, B-1050, Brussels, Belgium)

摘 要： 视觉特征提取是听视觉语音识别研究的热点问题。文章引入了一种稳健的基于 Visemic LDA 的口形动态特征，这种特征充分考虑了发音时口形轮廓的变化及视觉 Viseme 划分。文章同时提出了一种利用语音识别结果进行 LDA 训练数据自动标注的方法。这种方法免去了繁重的人工标注工作，避免了标注错误。实验表明，将 Visemic LDA 视觉特征引入到听视觉语音识别中，可以大大地提高噪声条件下语音识别系统的识别率；将这种视觉特征与多数据流 HMM 结合之后，在信噪比为 10dB 的强噪声情况下，识别率仍可以达到 80%以上。

关键词： 语音识别,听视觉语音识别,ASM, Linear Discriminant Analysis(LDA), Viseme

中图分类号： TP391.42 **文献标识码：** A **文章编号：** 1009-5896(2005)01-0064-05

A Robust Dynamic Mouth Feature Based on Visemic LDA for Audio Visual Speech Recognition

Xie Lei Fu Zhong-hua Jiang Dong-mei Zhao Rong-chun
Werner Verhelst* Hichem Sahli* Jan Conlenis*
(*School of Computer Science, Northwestern Polytechnical Univ., Xi'an* 710072, *China*)
*(*Dept of ETRO, Free University Brussels, Pleinlaan* 2, *B*-1050, *Brussels, Belgium*)

Abstract This paper presents a robust visual feature based on Visemic LDA for audio visual speech recognition, which captures dynamic lip contour information and reflects the viseme classes of visual speech. The paper also introduces an automatic labeling method using the speech recognition results for LDA training data, which avoids the tedious manually labeling work and labeling errors. Experimental results show that the audio visual speech recognition system based on the visual features presented in this paper can greatly increase the speech recognition rate in noisy conditions. The combination of the visual feature with multi-stream HMM can bring the recognition rate of over 80% at a 10dB SNR noisy condition.

Key words Speech recognition, Audio visual speech recognition, ASM, Linear Discriminant Analysis (LDA), Viseme

1 引言

当前语音识别技术面临的最大的挑战就是如何适应不同的应用环境，即如何消除噪声对识别造成的影响。噪声条件下语音识别稳健性研究主要包括识别器前端噪声滤除，建立包含噪声效果的改进的语音模型和提取稳健的语音信号特征。这些基于信号和模型的技术只能有限地提高识别率。

众所周知，人类在噪声环境下具有很强的语音感知和识别能力。这是因为人类语音感知是固有的多感觉道（Multi-modal）处理过程，包括对声学信号的全面分析以及高级别知识的获取，例如语法、语义和语用。当声学环境噪声存在时，一种具有重要作用的信息源就是读唇（Lipreading 或 Speechreading）。读唇是利用“视觉语音”，通过观察说话人的嘴唇动作来理解话意的方法。视觉语音主要通过三个方面来辅助语音感知[1]。首先它可以帮助听者进行声源定位，其次它包含有关语音的分割信息，再次它可以提供有关发音位置的信息，而发音位置的信息可以帮助分辨在声学空间中相近的发音。

视觉语音对听觉语音识别的辅助作用激发了对听视觉语音识别的研究，即如何提取集中于说话人面部（特别是嘴部）的视觉信息，并将其与常规的听觉信息相融合，从而为提高噪声条件下识别系统的稳健性开创了一条新途径。

视觉特征提取是听视觉语音识别的一个研究热点。目前的提取方法主要包括基于视频图像像素的提取方法和基于口形轮廓的方法[1]。前者对图像进行某种变换，将变换域的参数作为特征；后者则认为有用的视觉特征主要集中在口形的轮廓上，因此使用统计学模型来提取口形轮廓。基于像素的方法虽然可以保留牙齿、舌头的位置等视觉信息，但是容易受到视觉背景变化的影响，例如，光照强度变化和说话人的肤色都会影响所提取特征的“纯净性”；而基于轮廓的特

2003-07-11 收到，2003-11-25 改回
中国科技部与比利时弗拉芒大区科技合作项目(国科外字 1999(0209 号))资助课题

征不受视觉背景变化的影响，因而有可能带来更好的识别效果。

目前，基于轮廓的视觉特征主要是直接使用以帧为单位的口形轮廓点的坐标。这种方法虽然简单直接，但是没有考虑到语音的动态信息，然而语音的连续变化在视觉域上表现为口形的变化，因此本文在提取口形轮廓的基础上，引入了一种考虑口形动态变化的基于 Visemic LDA (Linear Discriminant Analysis) 的视觉特征，同时提出了一种利用语音识别结果进行训练数据自动标注的方法——语音识别 Viseme 标注法。文章内容安排如下：首先简单介绍口形轮廓的提取方法，然后着重讨论口形轮廓的动态特征及 Visemic LDA，接下来介绍本文所建立的听视觉语音识别系统以及识别实验分析，最后是结论。

2　口形轮廓提取

在提取视频动态特征之前，首先要准确地提取大量视频图像中的口形轮廓。由于嘴是一种柔性目标，因此其轮廓提取方法可以分为两类，即参数化方法和非参数化方法。参数化方法例如 ASM(Active Shape Model), AAM(Active Appearance Model)通过对标注好的训练数据进行统计分析，建立口形的模型；然后使用模型对新的口形图像进行轮廓匹配，从而得到口形轮廓。非参数化方法(例如 ACM, Active Contour Model）将口形轮廓提取看成是跟踪问题或者分割问题，无须进行训练，口形的轮廓信息可以从跟踪和分割的结果中获得。

ASM 是一种表述物体形状或轮廓的参数化统计模型，因此可以用于口形轮廓的提取。它通过对大量标定好的口形进行主分量分析（Principle Component Analysis, PCA），获得反映最大口形变化的若干个主轴，从而任何一个新的口形都可以映射到这些主轴上，即口形可以用这些主轴的加权和来表示。 视频图像序列中任何一个口形轮廓都可以通过对该模型的变形和匹配得到。口形轮廓一般用表征嘴唇外轮廓的 N 个坐标点表示，即

$$\boldsymbol{c}=(x_1,x_2,\cdots,x_N,y_1,y_2,\cdots,y_N) \tag{1}$$

3　基于 Visemic LDA 的口形动态特征

在得到口形轮廓之后，可以将表征轮廓的点的坐标直接作为视觉特征。但是，将这种视觉特征应用于听视觉语音识别的效果不是很理想，原因是没有考虑视觉语音的动态变化信息，而口形的变化反映了语音发音单元的分割信息，这对语音识别起到了重要作用。另外，听觉和视觉特征连接所组成的复合特征的维数过大（本文中，听觉 39 维＋视觉 32 维=71 维），使模型变得过于复杂。同时，表征轮廓的坐标点之间有可能存在信息冗余。因此需要在尽可能保留有用信息的同时，降低特征的维数。鉴于上述分析，为了使用低维特征反映视觉语音的动态变化信息，本文在提取的口形轮廓的基础上，引入基于 Visemic LDA 的视觉特征。整个视觉特征提取的流程如图 1 所示。

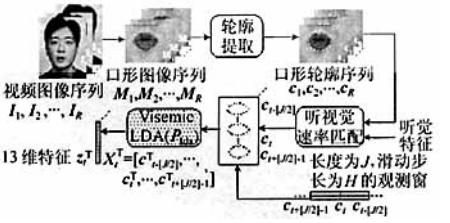

图 1　基于 Visemic LDA 的视觉特征提取过程

3.1 口形动态特征

由于视觉数据的采样速率（25Hz，即 25 帧/s）低于听觉特征的速率（100Hz），因此需要进行听视觉速率匹配（如图 1 所示），即将提取的口形轮廓进行插值，以得到相互匹配的听视觉数据速率（100Hz）。轮廓差值同时增加了训练数据的数量，有利于今后识别模型的参数估计。

为了反映视觉动态信息，使用一个长度为 J，步长为 H 的观测窗口在提取的轮廓序列 $\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_R$ 中滑动，将窗口内的 J 帧轮廓坐标连接起来，得到口形动态特征：

$$\boldsymbol{x}_t^{\mathrm{T}}=[\boldsymbol{c}_{t-[J/2]}^{\mathrm{T}},\cdots,\boldsymbol{c}_t^{\mathrm{T}},\cdots,\boldsymbol{c}_{t+[J/2]+1}^{\mathrm{T}}] \tag{2}$$

其中 $\boldsymbol{c}_t$ 是第 t 帧的轮廓坐标。

3.2 Visemic LDA

将轮廓坐标连接之后，得到的新的特征向量的维数很大（本文 32×9=288 维），因此需要降低特征的维数，同时保留最具有分辨性的信息。LDA [3]是一种经常用于分类和数据压缩的方法。对于 C 元分类问题（C 已知），如果用来进行训练的数据向量为 $\boldsymbol{X}_l, l=1,\cdots,L$，且已知每个训练数据向量所属的类 $\boldsymbol{x}_l\in c_i$，$i=1,2,\cdots,C$，则 LDA 的目的就是寻找一个投影矩阵 $\tilde{\boldsymbol{P}}$，使得投影后的训练数据（$\tilde{\boldsymbol{P}}\boldsymbol{x}_l, l=1,\cdots,L$）更好地分散到这 C 个类中。用公式表示，就是

$$\tilde{\boldsymbol{P}}=\arg\max_{\boldsymbol{P}}[\det(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{S}_B\boldsymbol{P})/\det(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{S}_W\boldsymbol{P})] \tag{3}$$

其中 $\boldsymbol{S}_B$ 和 $\boldsymbol{S}_W$ 分别表示训练数据的类间离散度（Between-class scatter）和类内离散度（Within-class scatter）。它们分别用

$$\boldsymbol{S}_B=\sum_{c_i\in C}\Pr(c_i)(\boldsymbol{m}^{(c_i)}-\boldsymbol{m})(\boldsymbol{m}^{(c_i)}-\boldsymbol{m})^{\mathrm{T}},\quad \boldsymbol{S}_W=\sum_{c_i\in C}\Pr(c_i)\boldsymbol{\Sigma}^{(c_i)} \tag{4}$$

来表示。其中 $\Pr(c_i)=L_{c_i}/L$，$c_i\in C$，称为经验概率分布函数；L_{c_i} 是 c_i 类训练样本的个数，L 是所有训练样本的个数；

$\boldsymbol{\Sigma}^{(c_i)}$ 和 $\boldsymbol{m}^{(c_i)}$ 分别是 c_i 类中训练数据的协方差阵和均值。所有训练数据的均值用式（5）来表示。

$$\boldsymbol{m}=\sum_{c_i\in C}\Pr(c_i)\boldsymbol{m}^{(c_i)} \tag{5}$$

为了得到投影矩阵 $\tilde{\boldsymbol{P}}$，计算矩阵对 $(\boldsymbol{S}_B,\boldsymbol{S}_W)$ 的广义特征值和右特征向量，从而满足 $\boldsymbol{S}_B\boldsymbol{F}=\boldsymbol{S}_W\boldsymbol{F}\boldsymbol{D}$。其中矩阵 $\boldsymbol{F}=[f_1,\cdots,f_d]$，其列向量 $\boldsymbol{f}$ 为广义特征向量；$\boldsymbol{D}$ 为对角阵，其对角元素为广义特征值，D 维最大的特征值排列在矩阵 $\boldsymbol{D}$ 中 $j_1,\cdots,j_D$ 的对角线位置上。数据 $\boldsymbol{x}_t$ 的新的 D 维特征向量可以用

$$z_t=\tilde{\boldsymbol{P}}\boldsymbol{x}_t=\boldsymbol{P}_{\mathrm{lda}}\boldsymbol{x}_t \tag{6}$$

计算出来，其中 $\boldsymbol{P}_{\mathrm{lda}}=[f_{j_1},\cdots,f_{j_D}]$。$f_{j_d}$（$d=1,\cdots,D$）叫做线性判别本征序列（Eigensequnece），它对应着数据投影之后产生高判别性的若干个方向。

对高维的动态口形特征进行 LDA 投影，需要首先确定分类的个数 C。我们知道，音素（Phoneme）是语音在听觉域的划分，它们通常作为识别模型的建模单元。然而不是所有的音素在视觉域上都是可分的，一个更好的选择是基于 Viseme 的划分。在这里，Viseme 是指一段语音所对应的特定口形序列。研究表明，音素在视觉域上可以聚类成若干个视觉 Viseme 类，即音素和 Viseme 之间存在一定的映射关系。如图 2 所示，44 个英语音素可以聚类为视觉域中的 13 个 Viseme（va，vb,⋯,vm）[4]，这 13 个 Viseme 在听觉域里又可以分为静音(A)，元音(B)、半元音(C)和辅音(D)。

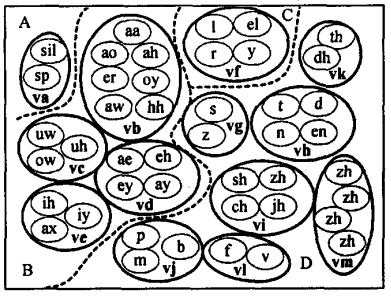

图2 44个英语音素在视觉域中聚类成13个Viseme

考虑到上述视觉语音的特点，假设所有的口形动态特征可以分配到这 13 个 Viseme 类中，使用式（6）对所有口形动态特征向量 $\boldsymbol{x}_t$ 进行 LDA 投影，在投影过程中只保留矩阵 $\boldsymbol{F}$ 的 13 个最大特征值所对应的特征向量，最终得到 13 维新的特征向量 z_t。我们把这种使用视觉语音 Viseme 划分的 LDA 投影叫做 Visemic LDA。它将口形动态特征映射到 13 个 Viseme 类中，从而使新的低维数的特征向量不仅包含了口形轮廓的动态变化信息，而且反映了对于视觉 Viseme 类的最大分辨信息。

3.3 语音识别 Viseme 标定法

在使用 Visemic LDA 进行特征投影之前，一项重要的工作就是对训练数据进行分类，将用于训练的视频图像序列进行 Viseme 标定。由于两个 Viseme 之间的边界是很难判定的，因此在对视频图像进行标定时，需要仔细观察口形的变化，这使标定工作变得非常繁重，且很容易出错。本文充分利用语音识别的音素级结果，引入了一种对训练数据进行自动标定的方法，称为语音识别 Viseme 标定法。由于语音识别系统在安静环境或者训练和识别条件匹配的情况下，具有相当高的识别率（例如本文后面所介绍的英文连接数字识别，在安静环境下的识别率高达 98%以上），因此，在得到语音识别的音素级划分结果之后，通过图 2 中的音素和 Viseme 映射关系，得到 Viseme 的划分（即标定）结果，如图 3 所示。这种语音识别 Viseme 标定法可以做到训练样本的自动分类，同时减少了手工标定错误的发生。

4 听视觉语音识别系统

本文所建立的听视觉语音识别系统使用 HTK3.1[5]构建。语音信号在进行预加重和加窗（窗长 30ms，步长 10ms）之后，提取 13 维 Mel 倒谱系数，同时为了捕捉听觉语音信号的动态信息，还结合了 Mel 倒谱系数的差分和二次差分，最终得到 39 维听觉特征。然后将听觉特征和经过 Visemic LDA 投影的视觉特征进行帧同步连接，产生听视觉复合特征。

一种实现听视觉语音识别系统最简单直接的方法是使用听视觉复合特征，为听觉和视觉流建立单一的 HMM 模型库和单一的识别器(本文称为 AVFC-VLDA, Audio Visual Feature Concatenate-Visemic LDA)。这种方法易于实现，但是没有考虑听觉视觉之间的相互影响和作用，因而达到的效果是不理想的。另外一种方法是使用如图 4 所示的多数据流 HMM（Multi-stream HMM）[6]。多数据流是指系统的输入来自多种数据源；多数据流 HMM 试图为不同的数据流独立建模，然后使用某种预先设定的同步点来合成模型。在听视觉语音识别系统中，输入数据由听觉流和视觉流组成。

在多数据流 HMM 中，假设听觉和视觉流在某一时刻 t 状态 j 产生各自观测向量（$\boldsymbol{O}_a$ 和 $\boldsymbol{O}_v$）的概率为 $P(\boldsymbol{O}_a^{(t)}\mid j)$ 和 $P(\boldsymbol{O}_v^{(t)}\mid j)$，则最终产生复合观测向量 $\boldsymbol{O}=[\boldsymbol{O}_a,\boldsymbol{O}_v]$ 的概率为

$$P(\boldsymbol{O}^{(t)}\mid j)=P(\boldsymbol{O}_a^{(t)}\mid j)^{\alpha}+P(\boldsymbol{O}_v^{(t)}\mid j)^{1-\alpha},\quad 0\le\alpha\le 1 \tag{7}$$

α 是数据流指数，它决定着听觉流和视觉流对输出概率的贡献。多数据流 HMM 考虑了听视觉流之间的相互作用，且可以通过数据流指数控制各个数据流的影响，因此非常适合于听视觉语音识别。

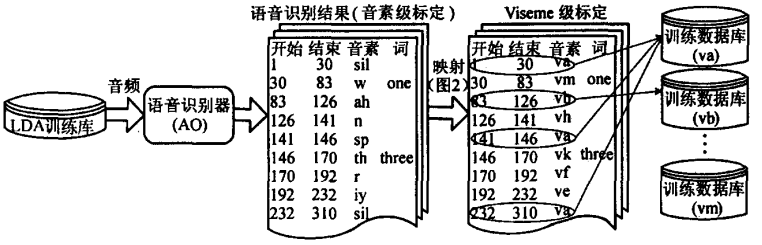

图 3　语音识别 Viseme 标定法

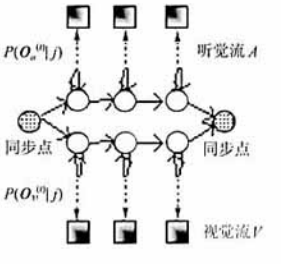

图 4　多数据流 HMM

5　识别实验及结果分析

本文所进行的听视觉语音识别实验是针对任意长度英文数字串（0～9）的识别。实验数据取自一个小型听视觉数据库 AVCONDIG。该数据库由机器随机产生 100 个连接数字串，在安静的实验室环境下（信噪比 30dB）使用数码摄像机记录一个说话人朗读这些数字串的面部视频和音频，产生 100 个音视频文件（AVI），共计 524s，13100 帧（25 帧/s）。

本文所涉及的口型轮廓采用 16 个典型坐标点来表示，如图 5 所示；视频图像中口形轮廓的提取采用一种改进的 ASM 算法：将所有发音所涉及的口形分为 3 类：闭合、张开和圆口。使用等量的训练数据为每类口形分别建立 ASM 模型，同时使用所有的训练数据建立一个全局 ASM 模型，然后使用这些 ASM 模型为每一个新的图像提取口形轮廓，使用判决准则选取最优的轮廓作为提取结果；最后根据说话时口形连续变化的特点，对轮廓进行平滑修正。这种方法只需要很少量的训练数据，就可以达到 98%以上的轮廓提取准确率，同时减少了手工标注训练数据和进行错误校正的工作。

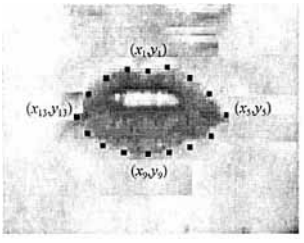

图 5　16 个坐标点所表示的口形轮廓

由于一个音素（或 Viseme）的平均持续时间约为 80～120ms，通过对比实验，最终选定长度 $J=9$（90ms），步长 $H=1$ 的观测窗口，用于动态口型特征的提取。尔后动态口形特征（32×9=288 维）再经过 Visemic LDA 投影，得到 13 维保留了口形动态信息的低维视觉特征。

为了进行噪声环境下语音识别实验，对 AVCONDIG 中的音频加入信噪比为 28dB，25dB，23dB，20dB，15dB，10dB 的鸡尾酒会噪声（Babble）。识别实验采用训练和识别条件不匹配的方式，即使用安静环境下录制的 85 个数字串进行 HMM 训练，15 个数字串（干净和加噪）作识别。实验中所建立的识别系统及其识别结果如表 1 所示。其中对于使用多数据流 HMM 的系统：AVMS（Audio Visual Multi-Stream），和 VLDA-VLDA(Audio Visual Multi-Stream-Visemic LDA)，数据流指数 α 的选择采用最小识别错误法，即在各种信噪比下尝试各种不同的值，选择使识别错误最小的值作为数据流指数。

从表 1 中可以看出，在安静条件下（30dB）常规听觉语音识别系统 AO（Audio Only，简称 AO）的识别率可以达到 98.48%。但是随着声学噪声的不断增加，识别率迅速下降。当信噪比达到 10dB 时，识别率只有 15.15%。两种视觉语音识别系统：VO(Visual Only)和 VO-VLDA (Visual Only-Visemic LDA)的识别率不受声学噪声的影响；使用口形轮廓直接作为视觉特征的 VO 系统的识别率可以达到 43.94%；口型轮廓经过 Visemic LDA 投影之后的视觉语音识别系统（Visual Only-Visemic LDA(简称 VO-VLDA)），其识别率比单纯使用口形轮廓作为特征的 VO 系统提高了 9 个百分点，达到了 53.03%。各种听视觉语音识别系统由于使用了不受声学噪声影响的视觉特征，因而使得噪声环境下的识别率有了不同程度的提高；在口形轮廓进行 Visemic LDA 投影之后，AVFC-VLDA 系统的识别率比 AVFC 系统平均高出 7.4 个百分点，比 AO 平均高出 17.8 个百分点；AVMS-VLDA 系统的识别率比 AVMS 系统平均高出 14.1 个百分点，比 AO 平均高出 29 个百分点。这说明，本文引入的这种反映口形动态信息和视觉 Viseme 特性的 Visemic LDA 投影特征是稳健的；尤其是将这种稳健特征和考虑听视觉之间相互影响的多数据流 HMM 相结合之后得到的 AVMS-VLDA 系统，其识别性能最优。即使在信噪比为 10dB 的强噪声环境下，识别率还保持在 80%以上。

另外，图 6 给出了使用多数据流 HMM 的系统在达到最高识别率时，信噪比（SNR）与数据流指数 α 之间的关系曲线。从曲线中可以看出，数据流指数 α 随着信噪比的逐步下降而下降。这说明随着噪声的增加，听觉流的可信赖程度逐步降低，视觉流的可信程度（$1-\alpha$）不断增加。当噪声很大时（SNR=10dB），识别系统在很大程度上依赖于视觉流，这也充分说明视觉信息对噪声条件下的语音识别起到了重要作用。

表1 实验识别系统及识别结果

系统	系统说明	特征维数	词识别率(%)						
			30dB	28dB	25dB	23dB	20dB	15dB	10dB
AO	常规听觉语音识别系统	39	98.48	80.30	77.27	65.15	51.52	42.42	15.15
VO	使用口形轮廓作为视觉特征的语音识别系统	32	43.94	43.94	43.94	43.94	43.94	43.94	43.94
VO-VLDA	口形轮廓经过 Visemic LDA 投影后的视觉语音识别系统	13	53.03	53.03	53.03	53.03	53.03	53.03	53.03
AVFC	听视觉特征进行帧同步连接的识别系统	39+32=71	86.36	81.82	80.30	77.27	74.24	59.09	43.94
AVFC-VLDA	Visemic LDA 视觉特征和听觉特征进行帧同步连接后的识别系统	13+39=52	96.97	86.36	81.82	79.92	76.13	68.18	65.56
AVMS	使用多数据流 HMM 的听视觉语音识别系统	39+32=71	86.36	83.33	81.82	78.79	74.24	68.18	62.12
AVMS-VLDA	使用 Visemic LDA 视觉特征的多数据流 HMM 识别系统	13+39=52	98.48	92.42	92.42	92.42	87.88	89.39	80.30

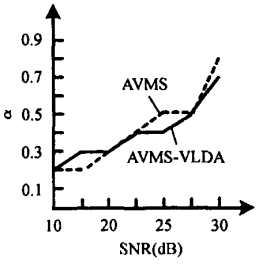

图6 信噪比与冥指数之间的关系

6 结论

本文在提取口形轮廓的基础上，引入了一种考虑口形动态变化和视觉 Viseme 划分的基于 Visemic LDA 的视觉特征；同时提出了一种充分利用语音识别结果进行 LDA 训练数据自动标定方法——语音识别 Viseme 标定法。语音识别实验表明，使用这种 Visemic LDA 视觉特征的听视觉语音识别系统的识别率要比仅仅使用口形轮廓特征的系统高出很多；尤其是使用 Visemic LDA 视觉特征与多数据流 HMM 相结合的系统，由于考虑了口形动态变化和听视觉流的信赖程度，在信噪比为 10dB 的强噪声环境下，识别率仍然能够达到 80% 以上，为噪声环境下的语音识别提供了可能。

参 考 文 献

[1] Potamianos G, Neti C, *et al*.. Recent advances in the automatic recognition of audiovisual speech. *Proc. IEEE*, 2003, 91(9): 1306 – 1326.

[2] Cootes T F, Taylor C J, *et al*.. Active shape models-their training and application. *Computer Vision and Image Understanding*, 1995, 12(1): 38 – 59.

[3] Neti C, Potamianos G, Luettin J, *et al*.. Audio visual speech recognition. Final Workshop 2000 Report, Baltimore, USA, 2000: 40 – 41.

[4] Rao C R. Linear Statistical Inference and Its Applications. New York, John Wiley and Sons, 1965: 122 – 128.

[5] Young S J, Kershaw D, Odell J, Woodland P. The HTK Book. http://htk.eng.cam.ac.uk/docs/docs.shtml, 2002.

[6] Dupont S, Luettin J. Audio-visual speech modeling for continuous speech recognition. *IEEE Trans. on Multimedia*, 2000, 2(3): 141 – 151.

谢 磊：男，1976年生，博士生，研究方向为语音信号处理.

付中华：男，1977年生，博士生，研究方向为语音信号处理.

蒋冬梅：女，1974年生，博士后，副教授，研究方向为语音信号处理.

赵荣椿：男，1937年生，教授，博士生导师，研究方向为语音图像处理与计算机视觉.

Werner Verhelst: 男，比利时布鲁塞尔自由大学教授，研究方向为语音信号处理.

Hichem Sahli：男，比利时布鲁塞尔自由大学教授，研究方向为数字图像处理.

Jan Conlines：男，比利时布鲁塞尔自由大学教授，研究方向为数字图像处理.