

# 基于 MMSE 谱减算法的农产品市场 信息语音识别技术

许金普<sup>1,2</sup>

(1. 青岛农业大学 动漫与传媒学院, 山东 青岛 266109; 2. 中国农业科学院 农业信息研究所, 北京 100081)

**摘要:** 为解决传统的便携式农产品市场信息采集设备操作不便,易受使用环境影响等问题,提出利用语音识别技术采集信息,以增加操作界面的灵活性。为增强语音识别的抗噪声鲁棒性,针对农产品市场信息采集的特殊工作环境,采集到 20 男 20 女语音训练集材料。首先利用最小均方误差 (MMSE) 谱减法进行前端带噪语音增强,得到增强后的语音信号,然后提取其 MFCC 特征用于 HMM 声学模型的训练;声学识别单元采用上下文相关的三音子模型,模型训练过程中采用了决策树状态聚类 and 增加高斯混合分量的策略,以提高模型的精确度。在 3 处不同环境不同信噪比情况下对训练出的模型进行测试,结果表明,MMSE 谱减算法处理后的语音识别率比基本谱减法 (SS)、多带谱减法 (MB) 有明显的提高,特别是在较低信噪比情况下更为明显。

**关键词:** 农产品市场信息; 语音识别; 谱减算法; 最小均方误差

**中图分类号:** S126      **文献标志码:** A      **文章编号:** 1004 - 3268 (2015) 05 - 0156 - 05

## Speech Recognition of Agricultural Market Information Based on MMSE Spectral Subtraction

XU Jinpu<sup>1,2</sup>

(1. College of Animation and Media, Qingdao Agricultural University, Qingdao 266109, China;  
2. Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

**Abstract:** In order to solve the inconvenient operation of traditional agricultural market information collection portable device, and other issues such as susceptibility to environment, we proposed the use of speech recognition technology to collect information in order to increase the flexibility of the operator interface. To enhance the robustness of speech recognition under the special working conditions of agricultural market information collection, we collected 20 males and 20 females voice training set material. Firstly, a noise spectral subtraction of the minimum mean square error (MMSE) at front-end was adopted, and then the MFCC features were extracted from enhanced speech signal for training HMM acoustic model; Acoustic identification unit was a context-sensitive Triphone model. In the training process, decision tree state clustering and increasing Gaussian mixture component strategy were adopted to improve the accuracy of the model. The trained models were tested in three different environments by different SNR speech sentences, and the results showed that the recognition rate of this method was more significantly improved than that of the basic spectral subtraction (SS), multi-band spectral subtraction (MB), especially at low SNR.

**Key words:** agricultural market information; speech recognition; spectral subtraction; minimum mean square error

农产品市场信息是农产品市场分析和预警的重要基础数据,是把握国内农产品市场变化趋势的重

要依据,对保证我国农产品安全、农业健康发展和农民持续增收都有重要意义。农产品市场信息关系到农业和社会稳定,是保证经济稳定健康发展的基础。针对农产品市场信息的重要性,国家相关的部门和地方政府也纷纷建立了各种形式的农产品市场信息采集平台,信息的采集往往利用传统的人工抄录再进行二次录入计算机、电话报价或邮件汇总等方式来完成<sup>[1]</sup>,但此类信息采集方法重复劳动较多,效率不高,且时效性较差。为此,很多机构和科研人员纷纷改进了信息采集的方法,开发了各种便携式的信息采集设备<sup>[2-6]</sup>。这些方式有自身的优点和便捷之处,提高了工作效率,但也存在不少问题。一般来说,便携式设备的屏幕和按键都较小,而农产品市场信息采集的工作场所往往是在室外,容易受到强烈光线、雨雾冰雪、恶劣天气、野外环境等条件的限制,给操作带来了不便<sup>[7]</sup>;另一方面,从人机交互的角度考虑,现有的信息采集设备主要依靠双手和视觉的配合来完成操作,但对需要人工干预情况下进行的信息采集,因其大部分是在生产过程、操作同时进行的,传统的设备必然导致操作人员暂时中断当前的工作转而进行信息的采集,这样就会导致劳动生产效率的降低。此外,传统的信息采集设备由于操作界面和提示信息复杂等因素,对操作人员的知识水平和认知能力都具有较高的要求。

近年来,随着语音识别(speech recognition)技术的迅速发展,基于语音交互界面的设备也在诸多行业开始应用。语音识别可以将语音转换为文本的形式,进而进行各种形式的处理和应用。从人机交互的角度看,语音交互界面是便携式设备最佳的人机交互接口<sup>[8]</sup>。经过几十年的研究,语音识别有了长足的发展。在相对安静的环境中,语音识别能取得非常好的识别结果。但在噪声环境下,现有的语音识别系统因受到噪声的污染,其识别性能则会急剧下降,尤其是在低信噪比的情况下,识别性能更为糟糕。造成这种低识别率的原因是识别环境与训练环境的不匹配<sup>[9-10]</sup>。噪声鲁棒性语音识别的研究目标就是消除或尽量减少这种不匹配现象,提高识别性能。而目前缺乏面向农产品市场信息采集领域的专用识别引擎,通用领域的语音识别系统往往是大词汇量连续语音识别,其模型数量庞大,对于便携式设备而言计算能力和存储空间都是无法接受的。因此,研究一种面向非特定人的、适用于农产品市场信息采集环境下的汉语普通话连续语音识别系统,利

用语音技术进行市场信息采集。

# 1 材料和方法

## 1.1 数据采集与预处理

所采用的语音库为自己录制的农产品价格信息句子发音,包括了 140 多种常见的鲜活农产品名称。句子的组成形式为“农产品名称 + 价格”,如“猪肉十三块二”“白菜一块三”等,句子是用剑桥大学的语音工具包 HTK 3.4 制作的转换语法生成,语法中的价格组合考虑了汉语的读音习惯,如 2.3 元一般读作“两块三”而不是“二块三”,0.5 元一般读作“五毛”等。转换文法部分如图 1 所示。

```
$digit_liang=一|两|三|四|五|六|七|八|九;  
$digit_1to9=一|二|三|四|五|六|七|八|九;  
$digit_2to9=二|三|四|五|六|七|八|九;  
$digit=一|两|三|四|五|六|七|八|九|十;  
$name=鹌鹑蛋|荸荠|扁豆|菠菜|菠萝|菜花|菜薹|蚕豆|草莓|橙子|大白菜|大葱|大蒜|冬瓜|豆芽|鹅|鸡蛋|番石榴|蜂蜜|佛手瓜|柑橘|橄榄|哈密瓜|海参|海带|海菜|海蜇|胡萝卜|瓠瓜|华莱士瓜|滑菇|黄瓜|黄花草|茴香|火龙果|鸡蛋|.....杨梅|杨桃|洋葱|椰子|伊丽莎白瓜|樱桃|油菜|柚子|原菇|枣|猪肉|竹笋|紫菜;  
(SENT-START $name (([$digit_2to9] + $digit_1to9) |  
([$digit_2to9] + [$digit_1to9] 块 $digit_1to9) |  
($digit 块 [$digit_1to9]) |  
($digit_liang 毛 [$digit_1to9]))  
SENT-END)
```

图 1 转换文法

根据上述转换文法生成一定数量的训练集和测试集录音脚本,指导录音。脚本文件可以转换为单音素和三音素的形式对声学观察序列进行标注,从而进行声学模型的训练及后续的识别。训练集录音地点为北京市的 2 处农贸市场,每处随机选择 10 男 10 女,每人朗读 140 个农产品名称和录音脚本中的 50 句话,共计录音  $(140 + 50) \times 40 = 7\,600$  句。测试集选择 3 男 3 女在相对安静的环境下录制,且测试集中的人员不在训练集中,每人朗读 50 句,然后分别混入农产品批发市场、社区小型农贸市场、超市环境下录制的噪声,各种噪声环境下分别得到信噪比(SNR)为 -5、0、5、10、15、20、25、30 dB 的测试语音。录音设备选用常见型号的手机,采样频率为 16 kHz,单声道录制,文件存储格式为 wav,量化精度为 16 bits。训练语音和测试语音最后经过人工准确地切分为句子。

## 1.2 试验设置

试验基于隐马尔科夫模型声学模型 HMM,声学特征参数是 39 维的 MFCC 特征,包括 12 维的静态特征和 1 维能量以及它们的一阶差分( $\Delta MFCC$ )和

二阶差分( $\Delta^2MFCC$ )系数。识别基元采用当前主流的声韵母三音子模型<sup>[9]</sup>,三音子模型考虑了上下文音节间的协同发音现象,能更好地反映音节间的过渡特征,比单音素、音节建模更稳定。HMM 模型的结构上采用自左向右的无跳转模型,其中第一个状态为起始状态,最后一个状态为终止状态,中间为输出状态。每个三音子模型有 5 个状态,静音模型‘sil’也包括 5 个状态,语音之间的停顿模型‘sp’包括 3 个状态,且与静音模型的中间状态绑定在一起。其他的一些声学参数设置为:分析帧长度 25 ms,帧间重叠 10 ms,加汉明窗,预加重系数为 0.97,滤波器组的个数为 26,倒谱加权滤波器的个数为 22。

1.3 最小均方误差 (MMSE) 语音增强方法

目前噪声鲁棒性语音识别研究中,在前端进行语音增强是一类重要的研究方法,而谱减类的语音增强算法实现简单,效果较好。其基本原理是从带噪信号中去除噪声信号,尽可能恢复纯净语音,以提高输入信号的信噪比。本研究采用基于短时幅度谱的 MMSE 估计方法<sup>[11]</sup>。得到纯净语音信号的估计幅度谱  $\hat{X}(\omega)$  有如下形式:

$$|\hat{X}(\omega)| = \left\{ \frac{\xi^p(\omega)}{\delta_p + \xi^p(\omega)} [ |Y(\omega)|^p - |\hat{D}(\omega)|^p ] \right\}^{1/p} \quad (1)$$

其中,  $\xi(\omega)$  是先验信噪比,  $|Y(\omega)|$  是带噪信号谱,  $|\hat{D}(\omega)|$  是估计的噪声谱。  $\delta_p$  对于给定的幂指数  $p$  取不同的常数。为防止出现负值,式(1)会采用谱下限。通过对衰减后的带噪语音谱  $\mu Y(\omega)$  ( $0 < \mu < 1$ ) 和前一帧增强且平滑后的语音谱进行平均,可以得到平滑后的频谱下限:

$$\mu |\bar{Y}(\omega)| = 0.5(\mu |Y(\omega)| + |\bar{X}_{prev}(\omega)|) \quad (2)$$

其中  $\mu |\bar{Y}(\omega)|$  表示平滑后的谱下限。即,如果式(1)增强后的频谱值小于  $\mu Y(\omega)$ , 则其取值为  $\mu |\bar{Y}(\omega)|$ , 最终纯净信号的估计幅度谱有如下形式:

$$|\hat{X}(\omega)| = \begin{cases} |\hat{X}(\omega)| & \text{如果 } |\hat{X}(\omega)| \geq \mu |Y(\omega)| \\ \mu |\bar{Y}(\omega)| & \text{其他} \end{cases} \quad (3)$$

$\xi(\omega)$  项对应于信号能量和噪声能量的比值,通

常称为先验(a priori)SNR。但由于无法得到纯净信号,因此该式不能直接计算,常用的一种近似的估计为<sup>[12]</sup>:

$$\xi(\omega) = (1 - \eta) \max \left( \frac{|Y(\omega)|^2}{|\hat{D}(\omega)|^2} - 1, 0 \right) + \eta \frac{|\hat{X}_{prev}(\omega)|^2}{|\hat{D}(\omega)|^2} \quad (4)$$

其中  $\eta$  为平滑常数,一般设为 0.98,式(4)是对当前瞬时信噪比(第 1 项)和过去信噪比(第 2 项)的加权求和。

2 结果与分析

试验分为模型训练阶段和测试阶段。在训练阶段首先利用本研究所述谱减法对带噪语音进行增强,获得去噪后的语音信号,然后提取 MFCC 特征,进行声学模型的训练。采用三音子识别基元模型数量会大大增加,对于有限的训练数据会产生训练不充分问题,因此训练过程中采用了决策树状态聚类方法,通过聚类可以使得具有相似发音特征的三音子共享模型参数,并且具有预测不可见三音子的能力<sup>[13]</sup>。同时,在聚类后统一将模型的高斯混合分量数目提升到 6 个,以增加模型的精度。测试阶段则不需要这 2 个过程,具体的流程如图 2 所示。

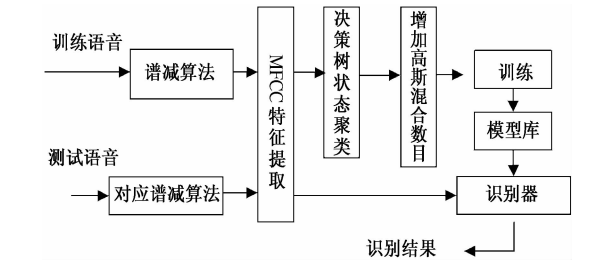


图 2 模型训练和语音测试流程

试验首先对加入大型农产品批发市场噪声的数据进行测试,分别对 -5 ~ 30 dB 的测试语音用基线系统(BaseLine)、基本谱减法(SS)<sup>[14]</sup>、多带谱减法(MB)<sup>[15]</sup>进行了测试,然后再用本研究提出的去噪算法进行测试。得出的词识别率如表 1 所示,其识别率曲线如图 3 所示。

表 1 不同算法在农产品批发市场的识别率

算法	SNR/dB								%
	-5	0	5	10	15	20	25	30	
BaseLine	18.04	25.27	45.36	67.96	84.93	90.94	91.78	92.69	
SS	20.70	39.95	70.02	85.31	88.89	90.72	90.41	88.28	
MB	19.10	35.01	62.79	81.89	88.66	91.48	91.63	92.24	
SMMSE	28.32	45.84	75.21	88.54	90.41	94.06	95.28	95.89	

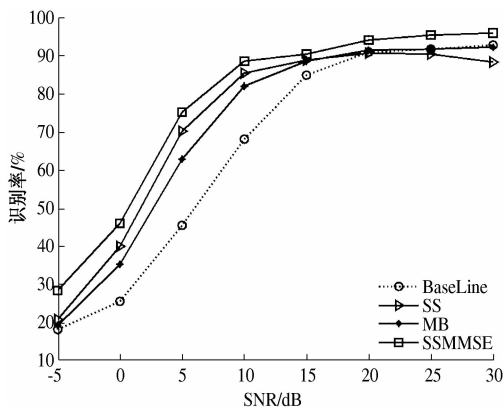


图3 大型农产品批发市场环境下不同算法的识别率

从表1和图3综合来看,采用语音增强方法后的识别率均比基线系统有了明显的提高,说明谱减法在语音识别前端进行处理是一种简单有效的方法。本研究提出的SSMSE算法性能优势较为明显,其次为SS和MB。与SS算法相比,SSMSE算法在低信噪比时的提升幅度较大,如在0 dB、5 dB、10 dB时识别率分别提高了5.89、5.19、3.23个百分点;而随着信噪比的提高,识别率提高程度不大。与MB算法相比,也出现类似的情况,如在0 dB、5 dB、10 dB时识别率分别提高了10.83、13.02、6.65个百分点,随着信噪比的提升二者性能差别不大。在社区农贸市场和超市环境下的试验也证实了本研究所提算法的有效性,其识别率曲线如图4和图5所示。

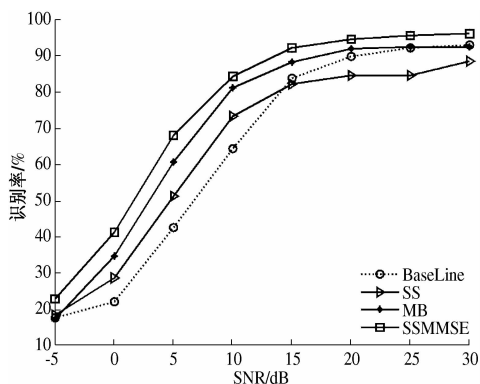


图4 社区农贸市场环境下不同算法的识别率

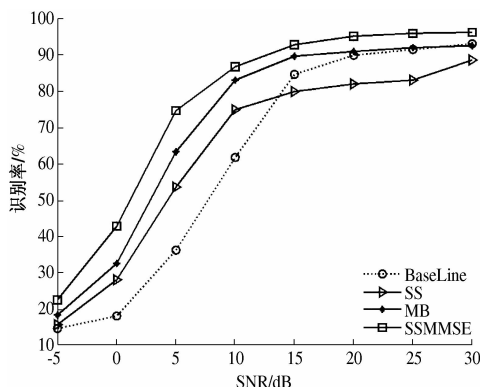


图5 超市环境下不同算法的识别率

值得注意的是,SS算法随着信噪比的提高( $>20$  dB),出现了识别率反而不及基线的情况,这说明该算法稳定性较差。而本研究所提算法则一直保持良好的稳定性,处在所有识别率曲线的最上端。

### 3 结论与讨论

本研究所提出的MMSE谱减算法,在最小均方误差意义下计算谱减参数,使得估计的语音信号与实际语音信号谱误差均方值最小。通过在大型农产品批发市场、社区农贸市场和超市等环境的试验表明,该方法能在前端有效地对输入语音信号进行增强,提高信噪比。与同类别的谱减算法相比有一定优势,特别是在0~15 dB时更为明显。

虽然本研究提出的算法能提高语音识别系统的识别率,在不同的环境下具有抗噪声鲁棒性的特点,但影响语音识别的因素很多,如口音、方言、说话方式、说话人生理状况等,这些因素不属于本研究的内容。本研究语音识别系统以略带口音的汉语普通话识别,且语音以朗读方式而不是以自然语言方式发音,并不适合方言识别。另外,农产品市场信息语音识别的词汇集合也是固定的,即根据已知训练集合的词汇量训练出相应的三音子识别模型,如果今后需要识别新的词汇,则需要重新训练。当前的农产品市场信息所要采集的词汇相对比较固定,主要包括名称、单价、等级、交易量、产地等,且语法构成方式相对简单,此类任务可以利用语音技术进行处理。

利用本研究提出的算法能将语音识别率提高,但从数据结果来看,在较低信噪比时的识别率依然较低,如在5 dB时仅为75.21%,0 dB时为45.84%,这仍旧有很大的提升空间。语音识别是一项复杂的任务,不是靠一两个算法能解决的,今后还需要结合其他的方法<sup>[16]</sup>,如特征空间、模型空间的算法,来进一步提高鲁棒性。

#### 参考文献:

- [1] 姚鑫,罗敏,杨国强.基于笔式交互的农产品信息采集系统的研究与设计[J].计算机与现代化,2012(4):71-75,78.
- [2] 邢振,郑文刚,申长军,等.农产品信息采集器:中国,CN202035021U[P].2011-11-09.
- [3] 李干琼,王东杰,于海鹏.“农信采”正能量[J].农产品市场周刊,2013(26):17.
- [4] 赵春江,申长军,邢振,等.农产品信息采集器及采集方法:中国,CN102122430A[P].2013-09-04.
- [5] 张石锐,郑文刚,申长军,等.基于GSM的农产品价格采集系统的设计与实现[J].农机化研究,2011(9):

193-196.

[6] 田文君,申长军,郑文刚,等.农产品价格信息采集与预警系统设计与实现[J].计算机工程与设计,2012(5):1816-1821.

[7] 许金普,诸叶平.一种基于语音识别的农产品信息采集设备和方法[J].农机化研究,2015(1):9-12, 17.

[8] 韩勇,须德,戴国忠.语音用户界面研究进展[J].计算机科学,2004(6):1-4,39.

[9] 倪崇嘉,刘文举,徐波.汉语大词汇量连续语音识别系统研究进展[J].中文信息学报,2009,23(1):112-123,128.

[10] 雷建军,杨震,刘刚,等.噪声鲁棒语音识别研究综述[J].计算机应用研究,2009,26(4):1210-1216.

[11] 雷建军,杨震,刘刚,等.基于短时谱估计的语音增强研究[J].计算机工程与应用,2008,44(32):6-9.

[12] Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator[J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1984, 32(6):1109-1121.

[13] 齐耀辉,潘复平,葛凤培,等.汉语连续语音识别系统中三音子模型的优化[J].计算机应用研究,2013,30(10):2920-2922.

[14] Boll S. Suppression of acoustic noise in speech using spectral subtraction [J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1979, 27(2):113-120.

[15] Kamath S, Loizou P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise [C]. International conference on acoustics, speech and signal processing, 2002:4164.

[16] 刘放军,王仁华.语音识别前端鲁棒性问题综述[J].计算机科学,2006(4):168-173.

\*\*\*\*\*

(上接第 155 页)

参考文献:

[1] Granado F, Olmedilla B, Blanco I. Serum depletion and bioavailability of lutein in type I diabetic patients[J]. European Journal of Nutrition, 2002, 41(2):47-53.

[2] 蒋水萍,张拯研,郑仕方,等.优化烟叶结构后不同采收成熟度对烤烟品质的影响[J].河南农业科学, 2013, 42(11):40-45.

[3] 黄彬城,季静,郑阳霞,等.植物类胡萝卜素的研究进展[J].天津农业科学,2006,12(2):13-17.

[4] Weeks W W. Chemistry of tobacco constituents influencing flavor and aroma [J]. Recent Advance of Tobacco Science, 1985, 11:175-200.

[5] 史宏志,刘国顺.烟草香味学[M].北京:中国农业出版社,1998:212-214.

[6] 周冀衡,杨虹琦,林桂华,等.不同烤烟产区烟叶中主要挥发性香气物质的研究[J].湖南农业大学学报:自然科学版,2004,30(1):20-23.

[7] 侯英,徐济仓,王保兴,等.叶黄素的热解产物分析[J].烟草科技,2007(12):27-32.

[8] 缪明明,王昆文,李鲜,等.叶黄素的化学降解产物及机理研究[J].烟草科技,1998(2):30-31.

[9] 刘金霞,李元实,姬小明,等.叶黄素氧化降解产物 GC-MS 分析及在卷烟加香中的应用[J].郑州轻工业学院学报:自然科学版,2011,26(2):24-27.

[10] 孟凡来,赵昶灵,段丽斌,等.高等植物类胡萝卜素的生物降解途径研究进展[J].中国农学通报,2013,29(24):143-150.

[11] 张永涛,刘惠芳,张东豫,等.类胡萝卜素的热裂解研究[C]//中国烟草学会工业专业委员会 2005 烟草化学学术研讨会论文集.海口:中国烟草学会,2005:335-339.

[12] 杨伟祖,谢刚,王保兴,等.烟草中  $\beta$ -胡萝卜素的裂解产物的研究[J].色谱,2006,24(6):611-614.

[13] 周芳芳,周丽娟,詹军,等.烟草功能菌的研究进展[J].河南农业科学,2013,42(12):6-10.

[14] Sanchez-Contreras A, Jimenez M, Sanchez S. Bioconversion of lutein to products with aroma[J]. Appl Microbiol Biotechnol, 2000, 54:528-534.

[15] Uenojo M, Marostica M R, Pastore G M. Carotenoids: Properties, applications and biotransformation in flavor compounds[J]. Química Nova, 2007, 30(2):616-622.